Professional Differences: A Comparative Study of Visualization Task Performance and Spatial Ability Across Disciplines

Kyle Wm. Hall, Anthony Kouroupis, Anastasia Bezerianos, Danielle Albers Szafir, and Christopher Collins

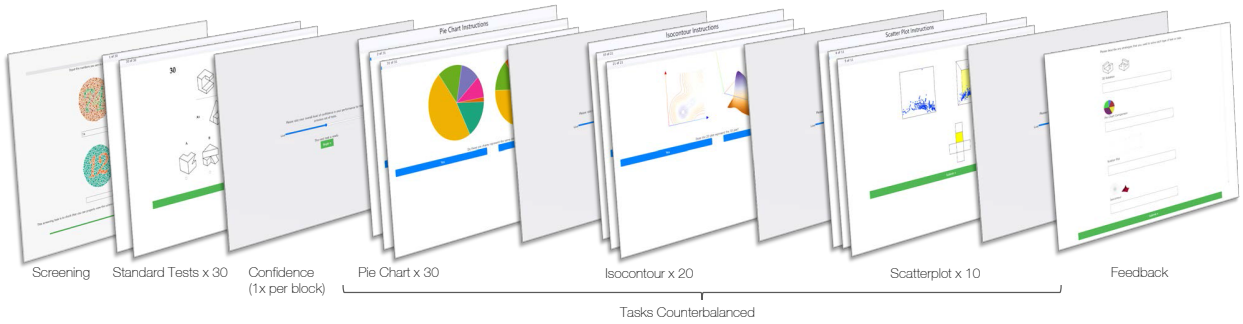

Fig. 1. We conducted a spatial abilities study across three professional populations. After screening, demographics, and a standardized set of spatial reasoning tasks, participants carried out three task blocks in counterbalanced order, followed by feedback on their problem-solving strategies.

Abstract—Problem-driven visualization work is rooted in deeply understanding the data, actors, processes, and workflows of a target domain. However, an individual's personality traits and cognitive abilities may also influence visualization use. Diverse user needs and abilities raise natural questions for specificity in visualization design: *Could individuals from different domains exhibit performance differences when using visualizations? Are any systematic variations related to their cognitive abilities?* This study bridges domain-specific perspectives on visualization design with those provided by cognition and perception. We measure variations in visualization task performance across chemistry, computer science, and education, and relate these differences to variations in spatial ability. We conducted an online study with over 60 domain experts consisting of tasks related to pie charts, isocontour plots, and 3D scatterplots, and grounded by a well-documented spatial ability test. Task performance (correctness) varied with profession across more complex visualizations (isocontour plots and scatterplots), but not pie charts, a comparatively common visualization. We found that correctness correlates with spatial ability, and the professions differ in terms of spatial ability. These results indicate that domains differ not only in the specifics of their data and tasks, but also in terms of how effectively their constituent members engage with visualizations and their cognitive traits. Analyzing participants' confidence and strategy comments suggests that focusing on performance neglects important nuances, such as differing approaches to engage with even common visualizations and potential skill transference. Our findings offer a fresh perspective on discipline-specific visualization with specific recommendations to help guide visualization design that celebrates the uniqueness of the disciplines and individuals we seek to serve.

Index Terms—visualization, spatial ability, perception, task performance, discipline, domain-specific, empirical evaluation

1 Introduction

Empowering individuals to understand their data is the enduring promise of the multidisciplinary venture that is visualization. While the approaches, materials, and philosophical perspectives of visualization continue to evolve, the goal of visualization remains largely the same: *how can we best serve individuals seeking to understand their data?* Visualization research has addressed this question following several lines on inquiry, including **1)** how people "see" data and **2)** how disciplines differ in terms of their data and tasks. The results of this work are foundational design guidelines grounded in the perceptions and reasoning processes of the monolithic "user" and discipline-specific characterizations of target problem spaces supporting nuances in design. Many have leveraged this pragmatic dichotomy to great success. Once we understand the data and discipline, we essentially rely on universal guidelines grounded in the generic "user" to create the right design.

While separating the discipline and data from the user simplifies design, prior studies have shown that reality is more complex: individuals have different cognitive abilities and those abilities inform visualization use [59, 72, 77]. Educators have explored in detail the relationships between cognitive abilities, educational path, and profession [61, 75, 76, 82]. However, we do not know how these two elements—profession and cognitive traits—interact when using visualizations. Should designers characterize the cognitive abilities of a target discipline as part of the design process, or is one-size-user-fits-all cognitive guidance about effective visualization designs sufficient?

We explore the relationship between profession, cognitive traits, and visualization use through the lens of variation in *spatial abilities*, defined as "the ability to generate, retain and manipulate abstract visual images" [41, p. 188]. Spatial abilities influence how effectively people use visualizations, including their strategies for exploring data and abilities to draw relations across that data [15, 19, 55, 72, 73]. However, prior training and experience can likewise influence an individual's abilities to use visualizations, even for people with low spatial abilities [27]. We

- *Kyle Wm. Hall was with Temple University during this work. Present Affiliation: Layer 6, E-mail: kyle@layer6.ai.*
- *Anthony Kouroupis and Christopher Collins are with Ontario Tech University. E-mail: anthony.kouroupis@ontariotechu.net, christopher.collins@ontariotechu.ca.*
- *Anastasia Bezerianos is LRI-Université Paris-Saclay. E-mail: anab@lri.fr.*
- *Danielle Szafir is with University of Colorado Boulder. E-mail: danielle.szafir@colorado.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

measured the associations between visualization performance, spatial ability, and discipline in an online study across computer scientists, chemists, and educators. Our results reveal that visualization task performance exhibits a substantive interplay between discipline and spatial ability, revealing limitations for the universal user approach employed by current visualization design processes. We provide a new lens for understanding visualization based on *professional cognitive differences* — systematic variations in disciplinary visualization usage concomitant with cognitive variations.

In this paper, we contribute: **1)** detailed statistical analyses demonstrating that visualization task performance (both correctness and completion times) varies with both discipline and spatial ability, **2)** a demonstration that enhanced correctness can coincide with other behavior (e.g., the perceived importance of graphical representations in the workplace), and **3)** a qualitative analysis of self-reported strategies, highlighting the importance of discipline-specific skills and concepts transferring to general tasks. Our results indicate that the expediency and efficiencies of prevailing disciplinary engagement and design models within visualization may have a hidden cost: designing for the universal user may obfuscate the full richness of whole persons whose professional choices and cognitive abilities are not easily separated.

2 Related Work

Our study bridges perspectives on visualization use across both problem-centered (i.e., domain expertise) and user-centered (i.e., user cognition) perspectives. We draw on prior knowledge and methods from across these fields to investigate the relationship between spatial abilities, professional discipline, and visualization comprehension.

2.1 Cognitive Visualization Studies

Visualization has a rich literature investigating psychological and cognitive factors associated with visualization use. This work typically focuses on understanding what elements of design make visualization generally useful for the typical user. However, this broad lens fails to consider nuanced differences emerging when any given person uses a visualization, such as ability levels [27, 71], prior knowledge [78], data literacy [48], and personality traits [56]. Moreover, such work is also counterbalanced by a growing body of work relating variations in visualization performance and usage to cognitive differences between individuals, otherwise known as *individual differences* [40]. In their framework characterizing individual differences for visualization, Peck et al. [59] proposed that cognitive differences between individuals could be classified according to three orthogonal dimensions:

- **Cognitive State**, which captures the instantaneous mental state of a person (e.g., in terms of cognitive load and emotions);
- **Cognitive Traits**, which are relatively stable over a person's adult life and include both personality traits (e.g., extroversion) and cognitive abilities (e.g., spatial ability); and
- **Experience/Bias**, which embodies the influence of a person's previous interactions and experiences.

By treating cognitive traits and experience/bias as orthogonal, this framework implicitly asserts that an individual's cognitive traits do not influence the experiences that individual pursues (e.g., education and profession) and vice versa. This decoupling is also present in visualization studies on individual differences where participant's professional information is typically either: **1)** not reported [34, 55, 72], or **2)** partially reported, but not analyzed as a potential factor influencing performance [14, 15, 19]. Studies [14, 15, 19] that do report professional information generally employ a professionally homogeneous participant pool, failing to consider the nature of professional cognitive differences and their relation to visualization. We next review work related to spatial ability, the cognitive trait we focus on in this study.

While domain-focused participatory methodologies imply that an individual's professional discipline can impact the right design for a given domain [28, 60], previous visualization work [15, 44, 73] has demonstrated that spatial ability has a complex influence on visualization use and preference. For example, Chen & Czerwinski [15] found that people's strategies for exploring visualized data change as a function of their spatial abilities. When navigating a 3D interface for semantic document relationships, recall was correlated with spatial ability whereas precision showed a negative correlation. Vicente et al. [73] found that people with low spatial ability took much longer to navigate hierarchical file interfaces and had different navigation patterns compared to subjects with high spatial ability. These performance differences persisted (at a reduced magnitude) for interfaces specifically designed to account for an individual's ability level [34]. Many studies of spatial ability emphasize people's abilities to make sense of more complex design choices, such as using 3D [72] or animation [19], finding that people with high spatial ability can generally make better use of such techniques. These differences may extend to different kinds of reasoning judgments: participants with higher spatial ability more easily use visualizations for Bayesian reasoning than individuals with lower spatial ability, and exhibit substantially greater correctness [55].

However, spatial ability is correlated with other factors that may affect visualization performance. The education literature demonstrates that there are substantial differences in spatial ability and other cognitive abilities between different professions (see 75 for a review). Moreover, VanderPlas & Hofmann [71] found that performance on line-up tests amongst undergraduate students is correlated with both visual ability and whether or not a student is a STEM major. Froese et al. [27] found that training can reduce differences between high and low ability people. Specific fields of training, duration of experience, and other factors associated with a profession, both at the undergraduate level [81] and beyond [75], may all similarly influence people's abilities to use visualizations. Given that different professions exhibit differing average spatial ability [75] and that performance on visualization tasks generally correlates with spatial ability, different professions may exhibit differing performance on basic visualization tasks. We test this conjecture and provide an initial exploration of cognitive differences between professions as they relate to visualization.

2.2 Visualization Design

The literature on visualization design (e.g., [26, 28, 36, 49, 52, 60, 62]) emphasizes exploring different domains to inform design processes, asserting that domains might differ in their data, processes and challenges, but not in terms of the cognitive traits of their constituent members. For example, Munzner's Nested Model [52] identifies key threats to the visualization design process and strategies to mitigate these threats. This model provides guidance for comprehending domain needs and evaluating how well a visualization addresses them; however, understanding the unique cognitive features of a target user group is not cast as a starting point for visualization design. Kirby and Meyer's characterization of visualization collaborations [36] explicates the critical role of cognitive and perceptual psychologists in system evaluation, but overlooks the potential for domain-specific cognitive and perceptual studies to formatively guide collaborative design. The reflective orientation of these models means that discovery of such differences might not occur until evaluating a fully-implemented design, potentially leading to missed opportunities and ineffective design choices.

To combat such hindsight discoveries, design frameworks typically require developers to explicitly characterize the needs and requirements of users early in the design process. Design Study Methodology [60] advocates that people understand the visualization literature pertaining to a problem before engaging in problem-driven visualization work, presumably including cognitive and perceptual work related to target data or tasks. However, this connection is limited to generalized perceptual knowledge and is distinct from discovering "the practices, needs, problems and requirements of the domain experts" [60, p. 2436]. The separation of user ability and domain knowledge neglects the possibility for the cognitive traits of domain experts to inform visualization design. Expanded methods [62] recommend employing liaisons (essentially domain-vis intermediaries) to translate domain processes and knowledge between visualization practitioners and domain users. While including liaisons may implicitly integrate some aspect of cognitive traits (e.g., differing capacities of individuals to use certain representations), liaisons do not account for the possibility of systematic cognitive differences between developer and domain.

The Activity Typology for Visual Analytics recommends developers characterize domain experts to enable "*Portable Analysis.* The ability to transfer analytic work across people, places, time, and devices" (emphasis in original text) [26, p. 271]. While cognitive variations between disciplines could limit such transferability, understanding the relationship between these variations and related visualization task performance between disciplines could also enrich design by helping characterize the disciplinary capabilities of target groups.

Recent frameworks provide more direct opportunities to investigate cognitive differences in formulating a design, although these opportunities are often implicit. For example, the *understand* activity within the Design Activity Framework [49] requires developers to understand a target domain and its users. While it emphasizes the role of tasks and workflows, the understand activity could include cognitive trait assessments across a discipline. Design by Immersion [28] offers a collaborative, transdisciplinary approach to problem-driven visualization in which a visualization practitioner's personal engagement, participation, and experience with a target domain informs the design. While Design by Immersion reflects on differences in knowledge, language and tasks, it does not capture potential cognitive differences between visualization practitioners and members of a target domain, instead offering the opportunity to implicitly assess such relationships through longitudinal collaboration with individual experts.

Explicitly accounting for cognitive traits in visualization design methodologies adds more layers of complexity to an already intricate process. While we have preliminary evidence for differences in cognitive abilities across disciplines [75, 81], for differences in visualization use across abilities [72], and for the role knowledge plays in shaping visualization use and interpretation [78], alternative work [57] suggests that data interpretation across disciplines may be more consistent than indicated by prior studies. We conducted a formal study across disciplines that conventionally reflect a range of spatial ability levels [75]—Education, Computer Science, and Chemistry—to investigate the interplay between disciplinary and cognitive factors in visualization interpretation in order to explore the need for understanding domain-specific cognitive traits in visualization design and evaluation.

3 Hypotheses

Inspired by the knowledge gaps identified in the previous work, and the potential importance of domain-level differences when designing and interpreting empirical evaluations of visualizations, we carried out a study of spatial abilities across domains of expertise.

For tractability, we focus our investigation on computer scientists, chemists, and educators. Computer scientists were an important population to include as they represent "standard participants" for many studies on task performances and design choices (e.g., ref. [14, 15]). In turn, performance differences between Computer Science and other disciplines could have broad implications (e.g., in terms of generalizability of previous work). We chose chemistry because: 1) there is the rich literature exploring relationships between spatial ability and chemistry performance (see ref. [30] and references therein), and 2) the research team had substantial experience working with chemists. Education was selected given the team's experience engaging with educators, and the critical role educators play in developing skills relevant to visualization (e.g., data literacy). Importantly, these disciplines, on average, reflect different levels of spatial ability. The physical sciences and computer science exhibit higher than average abilities, and educational professionals lower [75]. However, these broad categorizations neglect how both training and experience may shape visualization use and how individual abilities may vary within each discipline. Our study investigates three primary hypotheses about the relationship between disciplines, spatial abilities, and visualization use:

H1: Spatial abilities will differ across disciplines. Prior studies [61, 75] in educational psychology show that people's spatial abilities vary systematically across disciplines. Drawing on previous findings, we anticipate that chemists will have the highest average spatial abilities and educators the lowest. Measuring whether the relationship between spatial ability and professionalization results from systematic preferences (i.e., people of higher spatial abilities are drawn to chemistry) or are learned through training and experience is beyond the scope of this study. Nevertheless, a correlation between ability and discipline indicates fundamental differences in cognitive traits across disciplines that may impact visual reasoning processes.

H2: People's abilities to compare visualizations will correlate with their spatial abilities. People's search strategies and abilities to accomplish visual reasoning tasks, including how well they use particular kinds of visualizations, vary with spatial ability [15, 19, 55, 71–73]. People with high spatial abilities tend to use visualizations more effectively. We anticipate these patterns will replicate on an expanded set of visualization tasks, further emphasizing the importance of individual differences in design.

H3: People's abilities to compare visualizations will correlate with their disciplines such that average task performance will vary with discipline. Problem-oriented visualization design methods like design studies [60] or design-by-immersion [28] assume that disciplinary needs are core to the "right" solution to a visualization problem. We anticipate that these needs extend deeper than the data to encompass the abilities of the user directly. Discipline-specific training and experience can also build expertise and familiarity with particular visualization designs and tasks. For example, chemists have spent centuries experimenting with different strategies to represent molecular structures [21], and have explored techniques to represent multidimensional information on a 2D page, both molecular structures [53] and multidimensional functions [39]. Instruction on how to use chemical visualization permeates chemical education and corresponding texts [42]. When giving advice to potential authors, the editors at a prominent physical chemistry journal noted "well-composed and scientifically accurate figures constitute the core of a scientific paper" [35]. For disciplines like chemistry with a strong emphasis on visuals and systematized visualization education, we anticipate the interrelation of these factors and disciplinary differences in spatial abilities could lead to even higher performance. For example, it is reasonable to expect that chemists will perform better on tasks involving 2D projections of multi-dimensional datasets.

Confirming these hypotheses would indicate that effective visualizations need to consider not only the needs of the discipline but the abilities of the individuals within that discipline. Such results would reveal significant limitations in the idea of the "universal" user.

4 Study Design

To evaluate our hypotheses (Section 3), we measured: 1) spatial abilities using the Revised Purdue Spatial Visualization Test: Visualization of Rotations (PSVT:R), a psychometric instrument requiring participants to match different views on the same object through mental rotations [79, 80], and 2) people's ability to assess correspondences between different visualizations. Inspired by Just Noticeable Difference tasks from vision research [13], we selected the following visualization tasks where participants assess whether visualizations correspond to the same dataset at varying levels of dataset difference (difficulty).

Pie Chart Correspondence Can participants determine whether two pie charts contain the same data?

Isocontour Correspondence Can participants recognize whether a 2D isocontour projection matches a given 3D surface?

Scatterplot Correspondence Can participants assess which face of a 3D scatterplot corresponds a provided 2D projection?

The above visualization tasks share a common theme: can people reconcile different views of the same data? We chose this theme as it aligns with the emphasis on mental rotation in the PSVT:R and provides a complex task related ecologically valid scenarios (e.g., small multiples). Participants use complex visualizations to reason over high-level tasks that are relevant across disciplines. These tasks draw on elements of prior studies of visualization use and spatial abilities, such as search [15] and classification [71], extending the findings to a more complex suite of designs and broader set of individual characteristics. We chose to consider isocontour plots since chemistry (one of our target disciplines) often leverages these plots to discuss chemical processes [38], and chemistry students are trained to use them. Similarly, the 3D

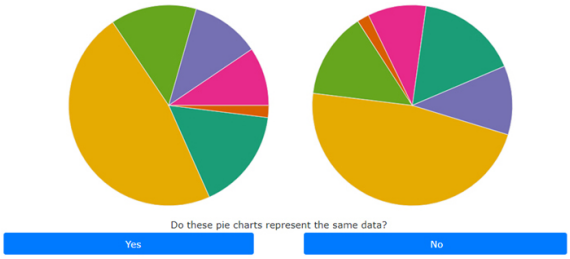

Fig. 2. Example of the Pie Chart Task. Participants completed 30 trials.

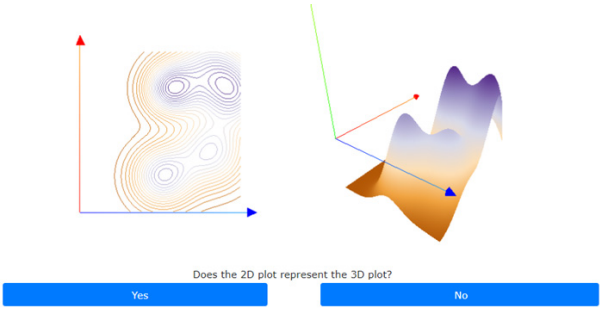

Fig. 3. Example of the Isocontour Task. The contour on the right starts axis-aligned and can be interactively rotated within a limited range. Participants completed 20 trials.

scatterplot task aligns with the common chemistry task of projecting 3D molecular structures to 2D representations [53]. We explored the above correspondences by creating a study for online deployment, consisting of a series of tasks for the participant to complete in sequence. The structure and setup of the study are detailed below; see Supplementary Material (SM) for snapshots of the study system. Our study code (stimulus generation, system, and analysis) and data are available on OSF [29].

4.1 Recruitment

We recruited participants from the different professions by circulating scripted recruitment emails on appropriate professional mailing lists and with institutional contacts (e.g., heads of departments), asking that they share the study with their faculty and graduate students. We also posted to relevant professional social media groups (e.g., LinkedIn, Facebook). In both cases, we informed people that they were free to forward the study to their contacts. We also leveraged our personal networks to initiate snowball sampling. To reduce variance in participant training, we restricted our recruitment to the USA and Canada. Participants were recruited to complete the tasks in a single session.

4.2 Study Structure

After participants agreed to the study consent form, the study process consisted of the following components, as depicted in Fig. 1. Participants confirmed readiness to begin the next block at each step, to allow for breaks as needed. There was a final screen where participants could provide information to receive compensation ($20 gift card). The study results and personal information required for compensation were stored in two separate unlinkable databases to preserve participant anonymity.

During piloting, it became clear that participants perceived the PSVT:R as the most difficult portion of the study, so we kept this test at the start of the study to ensure a more consistent experience across participants. Participants were provided with the standard instructions for the PSVT:R. For the Pie Chart, Isocontour, and Scatterplot tasks, participants were provided with instructions on how to complete each task. They were then presented with a training trial for that task, and provided feedback about their correctness on the training trail.

4.2.1 Measures

For each trial in all blocks, we collected the selected answer and completion time. Additionally, for the interactive tasks (Isocontour and Scatterplot), we also logged all mouse-based interactions in the study window. At the end of the PSVT:R and each task block, participants provided a self rating of their confidence on the previous block using a slider (Low–High) which was converted into a [0–100] scale.

4.2.2 Screening and Demographic Survey

Following the consent form, each participant completed an introductory survey to collect information about their profession, education history, visualization usage, participation in pastimes known to relate to spatial ability (e.g., video games), and demographic information. A second stage of screening for color vision acuity was carried out using four Ishihara plates [33], as several of our tasks relied on color discernment.

Only people who self-identified as a member of one of our target professions, had finished their undergraduate studies, and correctly answered all Ishihara plates, were permitted to continue the study. A failure to complete the plates correctly may be indicative of color-vision deficiencies or attributable to display conditions, so participants were not informed the reason for being screened out of the study.

We chose to exclude individuals who had yet to complete their undergraduate training as substantial shifts in discipline can take place during that time (e.g., people switching majors). Participants could only choose one discipline when selecting primary field (i.e., CS, Chemistry, OR Education), and our categorization of participant profession relies on this self-identification; we did not assign people to professional groups based on their education due to mapping challenges. We adopted this approach because our primary focus is assessing differences between disciplines (e.g., Education vs. Chemistry) rather than differences within individual disciplines (e.g., Educators). Exploring within-discipline variations could be especially relevant to Education where individuals often complete undergraduate training in one field and then specialize in education (see SM for a breakdown for our participants). For reference, we focused on recruiting participants with formal education training, and so "education" is thus a well-defined professional group for the purposes of this study.

We asked participants to complete the study using a computer, and excluded ones who attempted to complete the study on a mobile device, to increase consistency of display and interaction across participants.

4.2.3 PSVT:R

Participants then completed the PSVT:R [79, 80], a thoroughly researched spatial ability test [45–47, 79, 82]. The test is a standardized set of 30 rotation-based spatial ability tests in which participants match analogous rotations of 3D models. More specifically, participants are shown an initial view and and a subsequent rotated view of on asymmetric block. They are then provided with a second block, and must select from a set of distractors what the second block would look like if it were rotated the same way as the first. The test is image-based with no possibility of interaction.

4.2.4 Pie Chart Task

Each participant was sequentially shown 30 pairs of pie charts. All participants viewed the same 30 pairs of pie charts in random order. For each pair, participants had to decide whether the pie charts corresponded to the same dataset (*Do these pie charts represent the same data? Yes/No*, as seen in Fig. 2).

We created each pair of pie charts by randomly generating a pie chart consisting of 6 slices (random slice size of 10 degrees or greater, and random order), and then performing various perturbations to that pie chart (or lack thereof) to obtain the second pie chart for the stimulus pair while maintaining the first chart's color labelling. We used a consistent, approximately isoluminant color palette, avoiding visual pop-out effects between specific pairs of slices. The stimulus set contained five pie chart pairs for each of the below perturbation scenarios. For this discussion, *order* refers to the circular order of slices in a pie chart, and *orientation* is the angle a slice makes with the vertical direction. Rotating a pie chart changes its orientation, but not its order. We layered pie chart rotation and changes in slice order on top of data changes to capture scenarios that might more heavily rely on spatial ability; assessing data similarity in the presence of these additional non-data variations

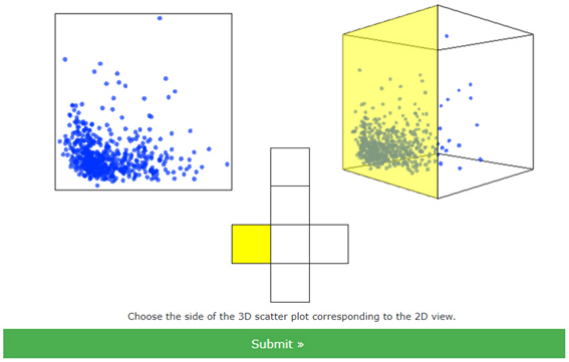

Fig. 4. Example of the Scatterplot Task. The 3D cube starts aligned to a random incorrect face and can be interactively rotated within a limited range. Linked highlighting supports face selection. Participants completed 10 trials.

involves mental rotations to a potentially great extent. Combining these factors together we arrived at six variants for the Pie Chart Task:

Identical Charts were completely identical, having the same slice sizes, orderings, and orientations.

Identical (Order Conserved, Rotated) Charts were identical, having the same slices sizes and ordering, but the second pie chart was a rotated version of the first chart.

Identical (Reordered) Charts were identical, having the same size slices, but the order of the slices was different between the two.

Different (Order and Orientation Conserved) Charts were different. The second pie chart had the same order and orientation of slices as the first pie chart, but two slice sizes had been altered.

Different (Order Conserved, Rotated) Charts were different. The second pie chart had the same order slices as the first pie chart but rotated, and two slice sizes had changed in size.

Different (Order Not Conserved, Rotated) Charts were different. The second pie chart did not have the same order of slices as the first chart, and two slice sizes were changed.

4.2.5 Isocontour Task

Each participant was sequentially shown 20 pairs of plots. Each pair consisted of one isocontour plot and one 3D surface plot, both of which used the same color ramp. All participants viewed the same isocontour-surface pairs, but in different random orders. For each pair, participants had to decide if the two charts corresponded to the same dataset (*Does the 2D plot represent the 3D plot? Yes/No*, as shown in Fig. 3).

For each 3D surface plot, we gave participants a randomly-generated initial viewing angle of the surface in profile at a slight acute angle with respect to the x-y plane. More specifically in spherical polar coordinates, the initial viewing angle consisted of: 1) a randomly-generated rotation around the vertical direction (i.e., the z axis) such that $\phi \in [0°, 360°)$, and 2) a second randomly-generated rotation from the z axis such that $\theta \in (70°, 110°)$. The rotations were performed with respect to the x-y centroid of each surface. Note that each participant saw each 3D surface plot from the same initial view. During each test (i.e., stimulus pair), participants were able to freely rotate the 3D surface about the vertical direction, and could rotate the viewing angle off of the x-y plane by $\pm 30°$, that is to say $\theta \in (60°, 120°)$. As a result, a participant could not precisely align the surface plot with the 2D isocontour, and mental rotation was in principle required to relate them.

We created each pair of plots by randomly generating a surface $z = f(x,y)$ composed of five Gaussian peaks on a flat plane (random both in peak position and size) for the isocontour plot, and then performed various perturbations to the initial surface (or not) to arrive at a second surface for the 3D surface plot. At the end of the generation procedure, all surfaces were normalized so that their vertical range was the same. The stimulus set contained five of each of the below types of isocontour-surface plot pairs. In addition to perturbing (or not) the positions of the peaks to arrive at the 3D surface, we also created some stimuli where the 3D surface was inverted (x,y,z) $\rightarrow$ (-x,-y,z) with respect to the isocontour surface, adding additional complexity and the potential for a false friend (a pair of plots that look potentially similar if one does not properly attend to the axes and their relation to gradations).

Identical The data sets were identical.

Different (Shifted Peaks) Two of the peaks in the underlying surface for the isocontour dataset were displaced to arrive at the dataset for the 3D surface plot.

Different (Flipped) The isocontour dataset was flipped along the y=x plane [i.e., (x,y,z) $\rightarrow$ (-x,-y,z)] to arrive at the dataset for the 3D surface, yielding two different surfaces with a symmetric relation.

Different (Flipped, Shifted) The isocontour dataset was flipped along the y=x plane [i.e., (x,y,z) $\rightarrow$ (-x,-y,z)] and then two of the peaks were shifted to arrive at the dataset for the 3D surface.

4.2.6 Scatterplot Task

Each participant was sequentially shown 10 pairs of 2D and 3D scatterplots. In each case, the 2D scatter plot corresponded to an orthographic view along one of the faces of the cube corresponding to the 3D scatterplot. For reference, the 3D datasets used for this task were generated by selecting sets of three dimensions from existing multi-dimensional datasets [25], and visually inspecting the datasets to ensure that the selection resulted in an asymmetric data distribution such that orthogonal views were visually distinct. The participant was tasked with selecting the 3D scatterplot face that corresponded to the 2D view (see Fig. 4). Importantly, a participant's initial viewing angle always aligned with the normal of a cube face other than that corresponding to the 2D scatterplot view, and the participant could only rotate around the face normal by 45°. As a result, a participant could not exhaustively search for the 2D view by rotating the cube, and mental rotation was in principle required to relate the 3D and 2D scatterplots. Note that the asymmetric nature of the datasets meant that the front and back sides of the 3D scatterplots were distinct, non-superimposable views.

The front face for a particular 3D scatterplot was the same for each participant, and the exclusion of the front face meant that the 2D view could only be one of the five remaining faces (though participants could select all six faces). In turn, the 2D view for each dataset was systematically varied across participants by professional group such that each of the five faces was equally balanced across that profession. The five possible face selections were also balanced within each participant with two datasets having 2D views corresponding to each possible face selection: top, bottom, left, right, and back ($2 \times 5 = 10$ pairs of 2D and 3D scatterplots). The dataset order was randomized per participant.

4.3 Strategies Questions

Previous work [66] has analyzed people's self-reported perspectives on their own practices to probe for discipline-specific approaches to tasks. Similarly, we asked participants to reflect on the strategies that they used when completing the PSVT:R test and each of the task blocks using free-form text. Thumbnail images were shown to remind participants of each task. We put this as the end of the study to avoid priming affects as participants went through the study. Note that the quality of the self-reported strategies likely depends on a participant's capacity for meta-cognition, and the extracted descriptions could be enhanced through observation and elicitation techniques. Such additional investigations were not amenable to our large online study.

4.4 System Details

We implemented a browser-based system in JavaScript. Pie charts, isocontour plots, and 3D surface plots were rendered with D3 [2], Three.js [5], and delaunator [3]. The Camera-Controls library [1] was used to provide click-and-drag functionality for rotating the 3D plots. Our system code is available on GitHub [4] and via OSF [29].

4.5 Participants

Due to the high educational expertise required for valid participants, we were only able to recruit 64 valid participants within the time-frame

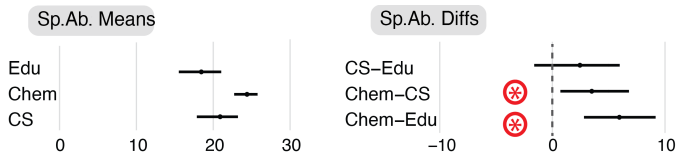

Fig. 5. *H1:* **Spatial Abilities** per Discipline, with the CI of means (left) and of mean differences (right). In the mean differences plot (right) CIs that are tighter and further away from 0 provide stronger evidence of differences. Red stars indicate evidence of differences between disciplines. Error bars represent 95% Bootstrap confidence intervals (CIs).

of the study. The number of participants varied per discipline: 33 Chemistry, 19 Computer Science (CS) and 27 Education. We note that even though our samples are not uniform, our results still provide valuable insights: there is no magic number for participants required in a study [7] and visualization studies often have small numbers of participants with relevant results [9, 10]. And when it comes to statistical evidence, confidence intervals (CIs) with just a hand-full of participants can still provide evidence of differences [24].

The gender statistics across the three disciplines were as follows. Chemistry: 58% female and 42% male. CS: 16% female, 74% male, 5% self-disclosed, and 5% not disclosed. Education: 81% female, 15% male, and 4% not disclosed. These are consistent with known professional gender differences in North America [54, 65]. The average age ($\pm$ standard deviation) for participants was 29$\pm$7 for Chemistry, 27$\pm$6 for Computer Science, and 37$\pm$9 for Education. We found no evidence of a difference in age for CS and Chemistry, but Education participants tended to be older; the details of these analyses are in SM, demographics supplement. Despite possible age differences, we note that cognitive traits are expected to be stable for adults [40, 59].

5 Results

We analyzed differences in performances and spatial abilities using sample means and 95% CIs, constructed using BCa bootstrapping (5000 iterations). We analyze the CIs using estimation techniques, interpreting them as providing different strength of evidence about the population mean, as recommended in recent reviews [8, 18, 22, 24]. CIs of mean differences that do not overlap 0 indicate a difference, corresponding to statistically significant results in traditional inferential statistics. However, CIs allow for more subtle interpretations: the farther from 0 and the tighter the CI is, the stronger the evidence. Equivalent p-values can be obtained from our CI results following Krzywinski and Altman [37]. All CIs are calculated using the R `boot` package [11, 23]. Here we report high level findings and mean/mean differences (detailed CIs in SM).

We analyzed differences in performance across spatial ability using a Pearson's correlation test. When $p < 0.05$ there is evidence of a correlation. We interpret correlation strength using the recommendations by Cohen [20], interpreting correlations as weak (R between 0.1–0.3), moderate (R between 0.3–0.5), and high (R between 0.5–1.0).

As our study was conducted online without supervision, we excluded as outliers participants whose accuracy in each task was below chance (below 50% for the Pie Chart and Isocontour, and below 16% for Scatterplots). We interpreted such behavior as a participant not understanding the tasks. We excluded five participants accordingly.

5.1 H1: Spatial Ability Per Discipline

To test if participants from different disciplines exhibit different levels of spatial abilities, we followed a procedure similar to Yoon et al. [81]. We calculated each participant's spatial ability as the number of correct trials in the PSVT:R rotation task (score out of 30). We then compared these spatial ability scores across our three discipline groups. Our results show (Fig. 5) that mean spatial ability was higher for Chemistry (24.4/30), followed by CS (20.9/30) and lowest for Education (18.5/30).

Looking at pair-wise differences of spatial ability, there are differences between Chemistry and the other two disciplines. There is strong evidence that spatial ability for Chemistry was higher than Education (by on average 5.9/30). Chemistry participants also had higher spatial ability than CS participants (by 3.5/30 on average). There may be a slight trend for CS participants to have on average higher spatial ability than Education participants, but evidence is not conclusive.

While not part of our main hypotheses, we also recorded participants' confidence in the PSVT:R rotation task. Chemistry and CS participants tended to be more confident than Education, but our evidence is not conclusive (see SM).

$\Rightarrow$ We partially confirmed **H1**, the spatial abilities of participants differed for some disciplines. The ranking of discipline in terms of spatial ability: Edu $\approx$ CS $<$ Chemistry.

5.2 H2 - Spatial Ability and Performance

To determine if spatial ability relates to performance, we looked at correlation between spatial ability and both time and correctness. We conducted our correlation analysis per task, as their complexity differs.

Correctness: When looking at each task separately (Fig. 6-Top), we see that for all tasks there is evidence ($p < 0.05$) of moderate to high correlation between spatial abilities and correctness for CS and Chemistry (all R 0.41–0.65). Higher spatial ability leads to higher correctness. For Education, we found a high correlation between spatial ability and correctness for the Isocontour and Scatterplot tasks (all R 0.53–0.61), and a trend for the Pie Chart.

Time: When looking at each task separately (Fig. 6-bottom), for Chemistry participants, there is evidence ($p < 0.05$) of moderate to high positive correlation between spatial abilities and time to complete the task for all tasks (all R 0.38–0.53). In other words, chemists with higher spatial abilities took longer than those with lower spatial abilities. We only saw a correlation between ability and time in Education for Pie Charts (again positive), and found no significant correlations with any tasks for CS.

$\Rightarrow$ When looking at performance overall, we partially confirmed **H2**: participants' correctness correlates to spatial ability (the higher the ability, the higher the correctness) for all tasks for Chemistry and CS participants, and for two of the three tasks for Education (Isocontour and Scatterplots, and there is even a trend for Pie charts). However, the correlation between spatial ability and completion times is less clear: for Chemistry participants, higher spatial ability correlates to higher time for all tasks, contrary to our hypothesis. We did not find a strong correlation between time and spatial ability for CS and Education.

5.3 H3: Performance Per Discipline

We hypothesized that performance (time, correctness) will differ depending on discipline, with Chemistry performing best and Education worst (following their spatial ability). We report next our analysis of time and correctness per task (seen in figures) and participants' self reported confidence (detailed CIs in SM).

5.3.1 Pie Charts (Fig. 7)

Correctness: Mean correctness in the Pie Chart task was higher for Chemistry participants (79.0%) followed by Education (76.4%) and then CS (75.6%). Looking at the mean differences, we found no evidence of a difference in correctness between disciplines.

Time: Mean times in the Pie Chart task was higher for Chemistry participants (9.8 s), followed by CS (9.4 s) and Education (7.9 s), with Looking at the mean differences, Chemistry was slower than Education (by 1.9 s) and CS slower than Education (by 1.4 s).

Confidence: Mean confidence was similar across participants for the Pie Chart task (63% for Chemistry and CS; 61% for Education).

5.3.2 Isocontour (Fig. 8)

Correctness: Mean correctness in the Isocontour task was again higher for Chemistry participants (76.8%), followed by CS (67.1%) and Education (60.9%). Looking at the mean differences, we found strong evidence of a difference in correctness between Chemistry and the other two disciplines—by 15.8% compared to Education and 9.7% compared to CS—but not between CS and Education.

Time: Mean times in the Isocontour task were again higher for Chemistry participants (87.9 s), followed by Cs (79.9 s) and Education

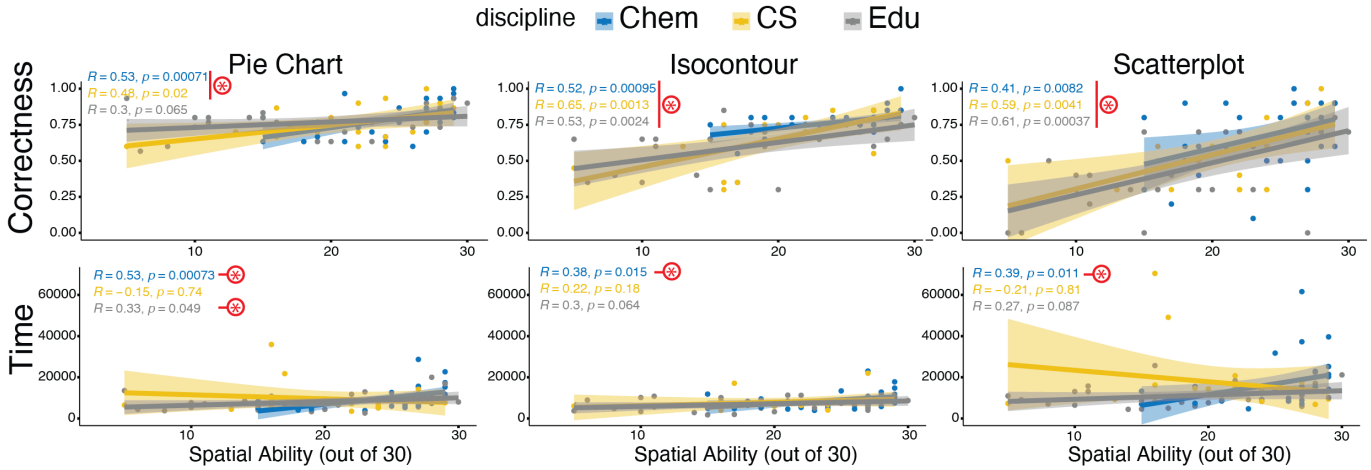

Fig. 6. *H2:* Spatial Abilities with respect to Performance: Pearson correlation between Spatial abilities and Correctness (top) and Time (bottom), divided by task and colored per Discipline. Red circles indicate evidence of correlation. Time in milliseconds.

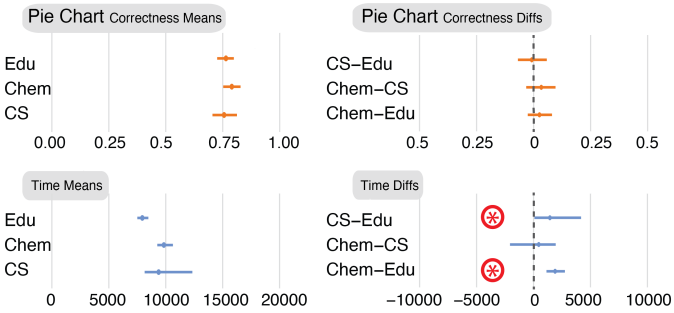

Fig. 7. *H3-Pie Chart* performance means (on the left) and mean differences (on the right), for Correctness and Time. Time in milliseconds.

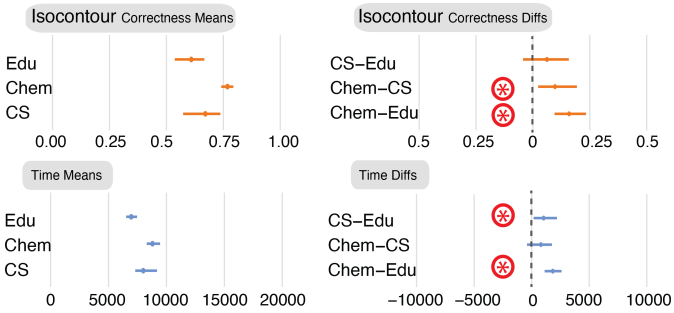

Fig. 8. *H3-Isocontour* performance means (on the left) and mean differences (on the right), for Correctness and Time. Time in milliseconds.

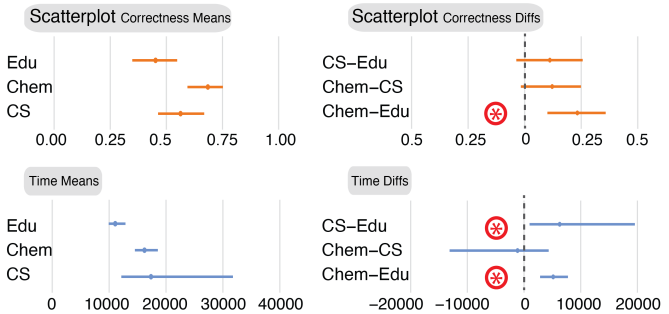

Fig. 9. *H3-Scatterplot* performance means (on the left) and mean differences (on the right), for Correctness and Time. Time in milliseconds.

(69.6 s). Looking at the mean differences, both Chemistry and CS were significantly slower than Education (by 1.8 s and 1.0 s respectively).

Confidence: Chemistry participants reported the highest confidence (74.8%), followed by CS (57.8%) and Education (38.2%). Looking at mean differences, Chemistry participants were more confident than both CS and Education (by 17.0% and 36.5% respectively), and CS participants were more confident than Education (by 19.5%).

5.3.3 Scatterplot (Fig. 9)

Correctness: The Scatterplot task was the most difficult overall. Mean correctness in the Scatterplot task was higher for Chemistry participants (68.4%), followed by CS (56.3%) and Education (45.2%). Looking at the mean differences, Chemistry was more accurate than Education by 23.3%. Chemistry trended towards being more correct than CS, and CS towards being more correct than Education, but our evidence is not conclusive.

Time: Mean times in the Scatterplot task was higher for CS participants (17.3 s), followed by Chemistry (16.2 s) and Education (11.0 s). Looking at the mean differences, both Chemistry and CS were slower than Education (by 5.1 s and 6.2 s respectively).

Confidence: For the Scatterplot task, Chemistry and CS participants reported similar confidence (52.9% and 53.4% respectively), with Education feeling the least confident (36.2%). Looking at mean differences, both Chemistry and CS participants were more confident than Education participants (by 16.6% and 17.1% respectively).

⇒ Our results partially confirm **H3**. Chemistry participants tend to perform differently than CS and Education. They are more correct than Education participants in two of three tasks (Isocontours and Scatterplots) and were more correct than CS participants for the Isocontour task and likely even for Scatterplots. Chemistry participants also tended to be slower than Education participants across all tasks.

The differences between CS and Education are more subtle. Although mean correctness was higher for CS, we found no substantial difference in correctness. But Education participants tended to be faster than CS participants in all tasks. Overall, we observed a time/error trade-off, with disciplines that are more correct (e.g., Chemistry) tending to also take more time to complete the tasks.

Self-reported confidence generally aligns with performance. Domains do not seem to differ in the Pie Chart, but we see Chemistry and CS participants being more confident than Education in the Isocontour and Scatterplot tasks. Chemistry participants are also more confident than CS for the Isocontour tasks, mirroring performance.

5.4 Strategies & Qualitative Analysis

Table 1 provides some highlights from behavioral questions that were asked at the start of the study. The final task of the study was to describe any strategies used for the PSVT:R and each task type. We tabulated all comments, separated by task type and professional group. Two members of the research team did a first pass open-coding on the comments to discover themes and common strategies. They then met to discuss the initial codes and develop a consensus code set. Finally,

Table 1. Summary of behavioral questions.

Characteristic	Chem.	CS	Ed.
Makes drawings or sketches for work	79%	68%	66%
Strongly Agrees with "Drawings and sketches are important to my work."	67%	37%	33%
Time spent playing video games			
0 h/week	27%	16%	41%
>5 h/week	24%	37%	7%
Time spent making visual art			
0 h/week	42%	37%	22%
>5 h/week	6%	5%	22%

Table 2. A selection of codes from task feedback analysis. Percentages represent participants coded at least once, by profession. The prefixes indicates relations to the three thematic groups in the main text: Strategies (s-), Generated Landmarks (g-), and Visual Landmarks (v-).

Code	Chem.	CS	Ed.
s-discrete steps	33%	58%	62%
s-skills transfer	12%	10%	0%
s-used prop or hand	21%	26%	50%
s-used symmetry	18%	0%	11%
s-used mirroring	21%	5%	0%
s-interaction	76%	58%	50%
s-had difficulties	6%	5%	27%
g-imagined axes	12%	0%	0%
g-imagined angles	24%	0%	8%
v-tracked unique features	36%	21%	15%
v-angles	30%	26%	11%
v-used visible axes	70%	63%	38%

one team member conducted a second pass coding on all data using the final code set. Codes were organized into three thematic groups:

Strategies (s) 13 sub-codes describing different problem-solving approaches (*repeated discrete steps*, *process of elimination*, *use of external aid*, *metaphor or explicit skill transfer*).

Generated Landmarks (g) 2 sub-codes describing mental projection of landmarks: *imagining reference axes* (in the PSVT:R) and *imagining reference angles* (mostly in the pie chart task).

Visual Landmarks (v) 9 sub-codes describing elements mentioned in the strategies (e.g., *angles*, *clusters*, *outliers*, *unique components*).

We rolled up the results across tasks to count participants who were associated with a code at least once. The codes which exhibited the greatest differences are summarized in Table 2. The complete data and a breakdown by task and profession provided in SM.

6 Discussion

Our results indicate that both an individual's cognitive abilities and discipline influence how they use visualizations. These differences manifest in both statistical performance and qualitative strategies. The relationship between discipline, abilities, and tasks offers several implications for visualization, including new insight into how we might reason about users in problem-driven design.

Spatial ability, discipline and performance We explored how both spatial ability and discipline influence task performance (correctness). While these two factors are often treated independently, our results reveal a complex interplay. We confirm past findings [75] that spatial ability varies with discipline (Fig. 5), with Chemistry having higher spatial ability than CS and Education (**H1**). We also found moderate linear correlations between task correctness and spatial ability across all three visualization tasks (Fig. 6-top, **H2**). These relationships were stronger in the Isocontour and Scatterplot tasks. One possible explanation for the greater correlation on these tasks is that they involve mental rotations of 3D visualizations, aligning more strongly with our spatial ability test. Chemistry substantially outperformed both CS and Education on the Isocontour task (see Fig. 8), and for the Scatterplot task outperforms Education and there is a trend towards outperforming CS. We failed to find notable differences between Education and CS, which aligns with the lack of difference in their spatial abilities (Fig. 5). In contrast, the Pie Chart task involves assessing 2D rotations of planar objects, and is thus more removed from our spatial ability test; here we found no differences in accuracy across disciplines (Fig. 7-top).

The fact that educators can exhibit differing abilities and performance is key for the visualization community. Teachers are society's knowledge stewards, responsible for guiding the development of regular citizens and future thought leaders alike. Supporting educators should be a top priority for visualization. Visualization literacy is critical for civic participation and leading-edge visualizations may miss wider adoption if they are inaccessible or ineffective for a key group that influences society's intellectual spheres from a formative stage.

Beyond correctness, we found disciplinary differences in completion times. In general, chemists took longer on tasks than educators (bottom of Figs. 7, 8, and 9), even on the pie chart task where their correctness was more comparable (Fig. 7-top). Furthermore, we found moderate positive correlations between completion times and and spatial ability across all three tasks for chemists: chemists with higher spatial ability were slower. We conjecture that this correlation might arise from individuals with higher spatial ability being able to assess the full complexity of the tasks. For example, in their free-form strategy descriptions, some chemists noted the mirror image relationship between opposite sides of the cube during the 3D scatterplot task whereas no computer scientists or educators noted that valid relationship. Critically, we must not confuse discipline capacity or difficulties with time. Designers may need to probe the multifacted processes of a discipline's visualization use, not reducing comparisons to single metrics.

Strategies and training Our survey results and participants' strategy comments provide nuanced insights into the performance results. For example, reflecting on the strategy codes (Table 2), all three disciplines described concrete strategies for tackling the Pie Chart task, in contrast to the Isocontour and Scatterplot tasks where the lower performance of educators coincided with vague strategies and comments about having difficulty with the task. Moreover, Pie Charts are relatively common visualizations, and included in many grade school curricula (thus, teachers are specifically trained to teach about them). This familiarity may explain the lack of substantial performance variations even though cognitive differences exist between the disciplines.

However, people reported different strategies across disciplines. For example, some chemists described comparing pie chart slices to characteristic angles like 90° in order to assess changes, while only two non-chemistry participant (both Education) reported using generated landmarks to complete any task. Knowledge of such strategy differences (even in the absence of performance differences) could provide a basis for engineering distinct techniques to further enhance each disciplines performance. The absence of performance differences for pie charts coinciding with strategy differences does raise additional questions: Do the strategy differences stem from training? Or do individuals develop different strategies, accounting for their cognitive differences, such that they can comparably engage with common visualizations?

We found explicit evidence that people's prior training can inform their visualization use via skill transference. Some chemists invoked concepts related to *chirality* (captured in code *s-symmetry* in Table 2) when describing how they worked through the PSVT:R test and Scatterplot task. Chirality captures the idea that molecules or other objects can be non-superimposable mirror images (e.g., like a person's left and right hands). Chirality and related concepts (e.g., handedness and stereochemisty) span synthetic chemistry [12, 42, 63] to biochemistry [74], and are an integral part of a chemistry education. The 3D blocks used in the PSVT:R and the scatterplots in our study are asymmetric (chiral) objects and previous education work [66] has demonstrated that chemists can transfer discipline-specific empirical strategies for addressing molecular problems to non-chemical abstract data and tasks. This skills transference aligns with chemists' improved performance on the PSVT:R and isocontour tasks. Such skills transference also

occurred at a lower level: one computer science participant noted using "intuition based on engineering design courses" when completing the PSVT:R test while other participants compared the isocontour task to navigating a mountain or a topographical map.

We also see evidence of skills transference in the higher use of props by the Education participants (double the other disciplines). Though gestures can improve spatial reasoning performance, people tend to use such gestures more frequently when they experience difficulties in assessing spatial rotation [17]. Educators are trained to teach students to approach problems from a variety of ways, including concretizing abstract problems [69]. This may point to the potential benefit of data physicalization for some people.

Skill transference has two possible consequences for design. First, we could look to existing interfaces and visualizations for a target discipline-specific task to guide tools that better encourage skill transference. Alternatively, if we know that a discipline performs better on a task using discipline-specific concepts, then that discipline's pedagogy may help visualization designers understand the visuals and metaphors used to introduce the concept and abstract these ideas to support and train other disciplines for whom a given task is difficult.

Disciplines and communication Participants' open-ended responses also illustrate potential communication barriers during problem-driven visualization design processes (e.g., design studies). Van Wijk [70] characterized the challenges during visualization collaborations in terms of knowledge and interest gaps including linguistic barriers (e.g., alternative understandings of the word "interphase"). Other work [28, 49, 60] stresses the importance of understanding and using a target discipline's language. While our tasks were discipline agnostic, the linguistic difference we observed across CS, Chemistry, and Education in describing their task completion strategies (see Table 2 and quotes in SM) highlight how individuals from different disciplines may describe a common task/experience differently. For example, on the PSVT:R test, many participants discretized the mental rotation process, but only chemists explicitly discussed imagined rotational axes (12%, *g-imagined axes* in Table 1). This linguistic difference may relate to discipline-specific training: rotational axes are integral to characterizing molecular symmetry, a key concept in chemistry [6, 31, 32]. However, our study involves no chemical/molecular information. Differences in language around core visualization concepts like tasks, when considered in conjunction with unique needs of a discipline, may pose unexpected barriers to understanding what people need to know about their data (e.g., task abstraction [49, 52, 60]). Design processes should pay close attention to how these differences can inhibit effective participatory design.

Demographic influences Variations in spatial ability and task performance are qualitatively mirrored by participant behaviours beyond discipline-specific skills (Table 1). Chemists are more likely to strongly agree that graphical representations of information are important to their work and to make drawings or sketches for their work. Educators are less likely to play video games compared to CS and Chemistry. If certain behaviors or attitudes coincide with spatial ability, then historical design approaches rooted in understanding a target discipline might indirectly capture some variations in cognitive ability. However, these types of connections are likely nuanced. For example, educators spent more time making visual art than CS or chemists (see Table 1), but exhibit lower spatial ability and task correctness scores.

Limitations & outlook The goal of our work was to take a first step in shedding light on professional cognitive differences and their implications for visualization. Therefore, we tested basic visualization skills, rather than embedding our study in difficult to control real-world tasks. While people do not regularly perform the exact study tasks, similarity assessments across visualizations are not unusual. For example, our pie chart comparison task maps to a small-multiples scenario where a person is assessing whether they have changed how they spend their days.

We cannot directly assess the causal origins of discipline variations observed in this study. Previous work [16, 43, 50, 64] indicates that spatial training regimes can improve spatial ability and discipline-specific performance/attainment. However, the gains from a particular training regime are not necessarily temporally enduring nor generalizable across multiple disciplines [50]. Furthermore, spatial ability interventions appear to have only a small effect on discipline performance [68], and longitudinal analysis [75] has demonstrated that individuals with high spatial ability are more likely to pursue careers in particular disciplines. It is an open question whether professional differences are due to training or the intrinsic capabilities of the people they attract.

It is worth noting that additional factors can influence task performance, such as representational fluency, which has been explored in the context of domain-specific problem solving [51, 67] and visualization [58]. However, we observed correlations between task performance and spatial abilities for all disciplines, so we anticipate that spatial ability is an important underlying source of variations between the disciplines. Natural differences in demographic factors exist across these populations (e.g., our participant groups exhibited differing levels of education attainment, gender balances, and average ages, see demographic SM). While they may contribute to the observed phenomena, several of these variations align with known differences between disciplines (e.g., gender-skewed male in computer science, but female in education). These differences were not controlled in our study and we thus cannot use them to explain our results. Nevertheless they open up intriguing avenues for further studies.

Regardless of the origins of the differences and other potential factors, our work illustrates it is not necessarily valid to treat different disciplines as equivalent in terms of task performance and spatial ability. Importantly, it is reasonable to conjecture that the observed differences are relatively robust, and that professional differences in task performance and cognitive abilities would endure, with implications for visualization. We did not recruit from individual institutions, but took a broad approach contacting groups of individuals across North America. Our study implicitly reflects a broad set of disciplinary training regimes.

In general, differences in cognitive ability, language, and familiarity with visualizations across disciplines all appear to affect how well people can use visualizations. However, the interplay between these factors is complex: neither ability nor discipline alone can explain the quantitative and qualitative differences in our data and further research is warranted. By studying visualization use across different disciplines and a range of cognitive abilities, we can begin to catalogue professional differences in cognitive abilities that drive more effective, discipline-specific visualization practices. We hope our work will encourage further research into the underexplored space of professional cognitive differences, looking to traits beyond spatial ability.

7 Conclusion

In this study, we explored visualization task performance as it related to both discipline and spatial ability, providing a new lens for understanding visualization use: professional cognitive difference. Our results provide preliminary evidence that critical differences exist between disciplines, and indicate the need to more deeply consider cognitive, social, and demographic factors in defining effective visualization. Our work also reveals that disciplines leverage distinct strategies and language to address discipline-agnostic visualization tasks, and these variations may or may not align with differences in visualization task performance and confidence. In general, disciplinary differences involve an intricate combination of social factors, training, and other individual differences that are likely interdependent. To fully empower individuals to understand their data, we must assess, respect, and celebrate such differences as we design visualizations. Investigating this rich space will require many additional studies and extensive future work, which we hope our results will inspire. Our study provides a first foray into this space, opening new avenues to better understanding the "user" in problem-driven visualization. Our data and code are available on OSF [29] to support and inspire future studies.

Acknowledgments

We acknowledge the support of NSF #1764089 & 2046725 and NSERC [RGPIN-2015-03916]. We thank Dr. So Yoon Yoon (University of Cincinnati) for providing the revised PSVT:R materials.

REFERENCES

[1] Camera-controls. Github: https://github.com/yomotsu/camera-controls.

[2] D3.js data-driven documents. Website: https://d3js.org/ Github: https://github.com/d3/d3.

[3] Delaunator. Github: https://github.com/mapbox/delaunator/.

[4] spatial-abilities-public. https://github.com/vialab/spatial-abilities-public.

[5] three.js JavaScript 3D library. Website: https://threejs.org/ Github: https://github.com/mrdoob/three.js/.

[6] P. Atkins and J. de Paula. *Physical Chemistry*. W. H. Freeman and Company, New York, USA, 8 ed., 2006.

[7] P. Bacchetti. Current sample size conventions: Flaws, harms, and alternatives. *BMC Medicine*, 8(17), Mar 2010. doi: 10.1186/1741-7015-8-17

[8] L. Besançon and P. Dragicevic. The significant difference between p-values and confidence intervals. In *Proc of the Conf. on l'Interaction Homme-Machine*, pp. 53–62, 2017.

[9] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg. The state of the art of spatial interfaces for 3D visualization. *Computer Graphics Forum*, 40(1):293–326, 2021. doi: 10.1111/cgf.14189

[10] K. Caine. Local standards for sample size at CHI. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pp. 981–992. ACM, 2016. doi: 10.1145/2858036.2858498

[11] A. Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2021. R package version 1.3-27.

[12] F. A. Carey and R. J. Sundberg. *Advanced Organic Chemistry Part A: Structure and Mechanisms*. Springer Science+Business Media, LLC, New York, USA, 5 ed., 2007.

[13] R. Chang, F. Yang, and M. Procopio. From vision science to data science: Applying perception to problems in big data. In *IS&T International Symposium on Electronic Imaging 2016 / Human Vision and Electronic Imaging 2016*, pp. HVEI–131. Society for Imaging Science and Technology, 2016.

[14] C. Chen. Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science*, 51(6):529–542, 2000. doi: 10.1002/(SICI)1097-4571(2000)51:6<529::AID-ASI5>3.0.CO;2-F

[15] C. Chen and M. Czerwinski. Spatial ability and visual navigation: an empirical study. *New Review of Hypermedia and Multimedia*, 3(1):67–89, 1997. doi: 10.1080/13614569708914684

[16] Y.-L. Cheng and K. S. Mix. Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, 15(1):2–11, 2014. doi: 10.1080/15248372.2012.725186

[17] M. Chu and S. Kita. The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140(1):102, 2011. doi: 10.1037/a0021790

[18] A. Cockburn, P. Dragicevic, L. Besançon, and C. Gutwin. Threats of a replication crisis in empirical computer science. *Communications of the ACM*, 63(8):70–79, 2020. doi: 10.1145/3360311

[19] C. A. Cohen and M. Hegarty. Individual differences in use of external visualisations to perform an internal visualisation task. *Applied Cognitive Psychology*, 21(6):701–711, 2007. doi: 10.1002/acp.1344

[20] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 2 ed., 1977.

[21] H. Cooke. A historical study of structures for communication of organic chemistry information prior to 1950. *Org. Biomol. Chem.*, 2:3179–3191, 2004. doi: 10.1039/B409980J

[22] G. Cumming. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, Taylor & Francis Group, New York, NY, 2012.

[23] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge, 1997.

[24] P. Dragicevic. Fair statistical communication in HCI. In J. Robertson and M. Kaptein, eds., *Modern Statistical Methods for HCI*, chap. 13, pp. 291–330. Springer, 2016.

[25] D. Dua and C. Graff. UCI machine learning repository. http://archive.ics.uci.edu/ml, 2020.

[26] D. Edge, N. H. Riche, J. Larson, and C. White. Beyond tasks: An activity typology for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):267–277, 2018. doi: 10.1109/TVCG.2017.2745180

[27] M.-E. Froese, M. Tory, G.-W. Evans, and K. Shrikhande. Evaluation of static and dynamic visualization training approaches for users with different spatial abilities. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2810–2817, 2013. doi: 10.1109/TVCG.2013.156

[28] K. W. Hall, A. J. Bradley, U. Hinrichs, S. Huron, J. Wood, C. Collins, and S. Carpendale. Design by immersion: A transdisciplinary approach to problem-driven visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):109–118, 2020. doi: 10.1109/TVCG.2019.2934790

[29] K. W. Hall, A. Kouroupis, A. Bezerianos, D. Albers Szafir, and C. Collins. Professional differences: A comparative study of visualization task performance and spatial ability across disciplines. `https://osf.io/shwzm/`. doi: 10.17605/OSF.IO/SHWZM

[30] M. Harle and M. Towns. A review of spatial ability literature, its connection to chemistry, and implications for instruction. *Journal of Chemical Education*, 88(3):351–360, 2011. doi: 10.1021/ed900003n

[31] J. M. Hollas. *Modern Spectroscopy*. John Wiley & Sons Ltd., Chichester, West Sussex, England, 4 ed., 2004.

[32] C. E. Housecroft and A. G. Sharpe. *Inorganic Chemistry*. Pearson Education Limited, Harlow, Essex, England, 3 ed., 2008.

[33] S. Ishihara. *Tests for Colour-Blindness*. Kanehara Shuppan, Tokyo, Japan, 1972.

[34] K. J.Vicente and R. C.Williges. Accommodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies*, 29(6):647–668, 1988. doi: 10.1016/S0020-7373(88)80072-5

[35] P. Kamat, G. V. Hartland, and G. C. Schatz. Graphical excellence. *The Journal of Physical Chemistry Letters*, 5(12):2118–2120, 2014. doi: 10.1021/jz500997e

[36] R. M. Kirby and M. Meyer. Visualization collaborations: What works and why. *IEEE Computer Graphics and Applications*, 33(6):82–88, 2013. doi: 10.1109/MCG.2013.101

[37] M. Krzywinski and N. Altman. Points of significance: Error bars. *Nature Methods*, 10:921–922, 2013. doi: 10.1038/nmeth.2659

[38] I. N. Levine. *Quantum Chemistry*. Pearson, New York, NY, 6 ed., 2008.

[39] I. N. Levine. *Physical Chemistry*, pp. 880–889. McGraw-Hill Higher Education, New York, NY, 6 ed., 2009.

[40] Z. Liu, R. J. Crouser, and A. Ottley. Survey on individual differences in visualization. *Computer Graphics Forum*, 39(3):693–712, 2020. doi: 10.1111/cgf.14033

[41] D. F. Lohman. *Spatial Ability: A Review and Reanalysis of the Correlational Literature*. School of Education, Stanford University, Palo Alto, CA, 1979.

[42] M. Louden. *Organic Chemistry*. Roberts and Company Publishers, Greenwood Village, Colorado, USA, 6 ed., 2009.

[43] T. Lowrie, T. Logan, and A. Ramful. Visuospatial training improves elementary students' mathematics performance. *British Journal of Educational Psychology*, 87(2):170–186, 2017. doi: 10.1111/bjep.12142

[44] W. Luo. User choice of interactive data visualization format: The effects of cognitive style and spatial ability. *Decision Support Systems*, 122:113061, 2019. doi: 10.1016/j.dss.2019.05.001

[45] Y. Maeda and S. Y. Yoon. A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualization tests: Visualization of rotations (PSVT:R). *Educational Psychology Review*, 25:69–95, 2013. doi: 10.1007/s10648-012-9215-x

[46] Y. Maeda and S. Y. Yoon. Are gender differences in spatial ability real or an artifact? evaluation of measurement invariance on the revised psvt:r. *Journal of Psychoeducational Assessment*, 34(4):397–403, 2016. doi: 10.1177/0734282915609843

[47] Y. Maeda, S. Y. Yoon, K. Kim-Kang, and P. K. Imbrie. Psychometric properties of the revised PSVT:R for measuring the first year engineering students' spatial ability. *International Journal of Engineering Education*, 29:763–776, 2013.

[48] H. Mansoor and L. Harrison. Data visualization literacy and visualization biases: Cases for merging parallel threads. In *Cognitive Biases in Visualizations*, pp. 87–96. Springer, 2018.

[49] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200, 2014. doi: 10.1109/TVCG.2014.2346331

[50] D. I. Miller and D. F. Halpern. Can spatial training improve long-term outcomes for gifted stem undergraduates? *Learning and Individual Differences*, 26:141–152, 2013. doi: 10.1016/j.lindif.2012.03.012

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

[52] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928,

2009. doi: 10.1109/TVCG.2009.111

[53] M. S. Newman. A notation for the study of certain stereochemical problems. *Journal of Chemical Education*, 32(7):344, 1955. doi: 10.1021/ed032p344

[54] OECD. Gender imbalances in the teaching profession. *Education Indicators in Focus*, 49, 2017. doi: https://doi.org/10.1787/54f0ef95-en

[55] A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. J. Han, and R. Chang. Improving bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):529–538, 2016. doi: 10.1109/TVCG.2015.2467758

[56] A. Ottley, H. Yang, and R. Chang. Personality as a predictor of user strategy: How locus of control affects search strategies on tree visualizations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3251–3254, 2015. doi: 10.1145/2702123.2702590

[57] L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):1–25, 2018. doi: 10.1186/s41235-018-0120-9

[58] P. Parsons. Promoting representational fluency for cognitive bias mitigation in information visualization. In G. Ellis, ed., *Cognitive Biases in Visualizations*, pp. 137–147. Springer Nature Switzerland, Cham, Switzerland, 2018. doi: 10.1007/978-3-319-95831-6_10

[59] E. M. Peck, B. F. Yuksel, L. Harrison, A. Ottley, and R. Chang. Towards a 3-dimensional model of individual cognitive differences: Position paper. BELIV '12. ACM, 2012. doi: 10.1145/2442576.2442582

[60] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012. doi: 10.1109/TVCG.2012.213

[61] D. Shea, D. Lubinski, and C. Benbow. Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93:604–614, 09 2001. doi: 10.1037/0022-0663.93.3.604

[62] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair. Bridging the gap of domain and visualization experts with a liaison. In E. Bertini, J. Kennedy, and E. Puppo, eds., *Proc. Eurographics Conf. on Visualization (EuroVis) - Short Papers*, pp. 127–131. Eurographics, 2015. doi: 10.2312/eurovisshort.20151137

[63] T. W. G. Solomons and C. B. Fryhle. *Organic Chemistry*. John Wiley & Sons Ltd., Hoboken, NJ, USA, 9 ed., 2008.

[64] S. Sorby, B. Casey, N. Veurink, and A. Dulaney. The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, 26:20–29, 2013. doi: 10.1016/j.lindif.2013.03.010

[65] Statistics Canada. Table 37-10-0018-01 postsecondary enrolments, by registration status, institution type, status of student in canada and gender, 2020. doi: 10.25318/3710001801-eng

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

[67] M. Stieff, M. Hegarty, and G. Deslongchamps. Identifying representational competence with multi-representational displays. *Cognition and Instruction*, 29(1):123–145, 2011. doi: 10.1080/07370008.2010.507318

[68] M. Stieff and D. Uttal. How much can spatial training improve stem achievement? *Educational Psychology Review*, 27(4):607–615, 2015. doi: 10.1007/s10648-015-9304-8

[69] J. Szendrei. Concrete materials in the classroom. In *International Handbook of Mathematics Education*, pp. 411–434. Springer Science+Business Media, Dordrecht, Netherlands, 1996.

[70] J. J. Van Wijk. Bridging the gaps. *IEEE Computer Graphics and Applications*, 26(6):6–9, 2006. doi: 10.1109/MCG.2006.120

[71] S. VanderPlas and H. Hofmann. Spatial reasoning and data displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):459–468, 2016. doi: 10.1109/TVCG.2015.2469125

[72] M. C. Velez, D. Silver, and M. Tremaine. Understanding visualization through spatial ability differences. In *Proc. IEEE Conf. on Visualization*, pp. 511–518, 2005. doi: 10.1109/VISUAL.2005.1532836

[73] K. J. Vicente, B. C. Hayes, and R. C. Williges. Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors*, 29(3):349–359, 1987. doi: 10.1177/001872088702900308

[74] D. Voet, J. G. Voet, and C. W. Pratt. *Fundamentals of Biochemistry: Life at the Molecular Level*. John Wiley & Sons Ltd., Hoboken, NJ, USA, 3 ed., 2008.

[75] J. Wai, D. Lubinski, and C. P. Benbow. Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4):817–835, 2009. doi: 10.1037/a0016127

[76] R. M. Webb, D. Lubinski, and C. P. Benbow. Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, 99:397–420, 2007. doi: 10.1037/0022-0663.99.2.397

[77] K. Wu, E. Petersen, T. Ahmad, D. Burlinson, E. S. Tanis, and D. A. Szafir. Understanding data accessibility for people with intellectual and developmental disabilities. *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, 2021. doi: 10.1145/3411764.3445743

[78] C. Xiong, L. Van Weelden, and S. Franconeri. The curse of knowledge in visual data communication. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):3051–3062, 2019. doi: 10.1109/TVCG.2019.2917689

[79] S. Y. Yoon. *Psychometric properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (The Revised PSVT:R)*. PhD thesis, Purdue University, 2011.

[80] S. Y. Yoon. Revised purdue spatial visualization test: Visualization of rotations (PSVT:R). [Psychometric instrument], 2011.

[81] S. Y. Yoon and E. L. Mann. Exploring the spatial ability of undergraduate students: association with gender, STEM majors, and gifted program membership. *Gifted Child Quarterly*, 61(4):313–327, 2017. doi: 10.1177/0016986217722614

[82] S. Y. Yoon and K.-H. Min. College students' performance in an introductory atmospheric science course: associations with spatial ability. *Meteorological Applications*, 23(3):409–419, 2016. doi: 10.1002/met.1565