MIND, BRAIN, AND EDUCATION

Fixation Disparity: A Possible Index of Visuospatial Cognition during Authentic Learning Tasks

Do Hyong Koh[1], Zhen Xu[1], Jiahui Wang[2], Andrea N. Burgess[3], Amanda Seccia[4], Matthew Schneps[5], Marc Pomplun[5], Richard Lamb[6], Andreas Keil[7], Kara Dawson[1], and Pavlo Antonenko[1]

ABSTRACT— This article describes a novel method for quantifying fixation disparity and evaluates its role in visuospatial cognition during an authentic learning task, specifically, the determination of molecule chirality in organic chemistry involving mental rotation and pattern comparison. The first study examined the influence of molecular model dimensionality (2D vs. 3D) on chirality determination performance and visual attention of 55 participants. The second study explored how the sustained playing of the tile-matching game Mahjong, a pattern comparison game, can affect visual attention and visuospatial performance during the chirality determination task of 59 participants. Fixation disparity was one of the eye tracking variables explored. Both studies revealed that (1) individuals with higher fixation disparity underperformed on the chirality task, which involves mental rotation and pattern comparison, and (2) fixation disparity improved over time in participants who played Mahjong. This work provides important implications for using fixation disparity as a possible biomarker of visuospatial performance.

[1]School of Teaching and Learning, College of Education, University of Florida
[2]School of Teaching, Learning and Curriculum Studies, Kent State University
[3]Department of Special Education, Vanderbilt University
[4]Department of Psychology, University of Chicago
[5]Department of Computer Science, University of Massachusetts Boston
[6]Department of Special Education, East Carolina University
[7]Department of Psychology and Center for the Study of Emotion and Attention, University of Florida

Address correspondence to Do Hyong Koh, College of Education, University of Florida, PO Box 117042, Gainesville, FL 32611-7044; e-mail: dohyong.koh@coe.ufl.edu

Visuospatial skills that underlie attention, cognition, and learning are very important in most domains—from the science, technology, engineering, and mathematics (STEM) disciplines of chemistry, physics, biology, geology nanoscience, mechanical engineering to visual and performing arts, design, construction, health sciences, and sport management (Atit et al., 2021; Castro-Alonso, 2019; Ehmann et al., 2021). As increasing evidence for the role of visuospatial skills in reasoning and problem-solving across domains accumulates, the belief that research into visuospatial cognition in educational settings and its relationship to student engagement, achievement, and persistence is becoming more widespread among teachers, scholars, and the public (Knapp, 2011; Lubinski, 2010; Uttal et al., 2013).

A host of cognitive factors, including differences in processing strategies, working memory capacity, and visual attention span, as well as socially mediated factors such as outcomes expectations, prior formal and informal learning experiences, and societal stereotypes, are thought to contribute to differences in visuospatial reasoning and performance (Atit et al., 2021; Halpern & Collaer, 2005). Much of the empirical research into the role of these variables and mechanisms have been conducted using measures of individual differences in visuospatial working memory (Allen, Higgins, & Adams, 2019), mental rotation skill (Alvarez-Vargas, Abad, & Pruden, 2020), and other visuospatial skills such as spatial navigation (Munion, Stefanucci, Rovira, Squire, & Hendricks, 2019). These individual differences are used to explain differences in participants' performance on visuospatial tasks, which have been explored objectively and directly using psychophysiological tools such as eye tracking. For example, several studies have employed an eye fixation measure (time looking at relevant information) and helped determine that (1) a strong link exists between eye fixations and learning outcomes (Mayer, 2010); (2) visual cues guide learners' visual attention (Boucheix & Lowe, 2010; De Koning, Tabbers, Rikers, & Paas, 2010); (3) prior knowledge guides visual attention (Canham & Hegarty, 2010; Jarodzka, Scheiter, Gerjets, & Van Gog, 2010); and (4) learners who view animation and on-screen text must split their attention between graphics and printed words (Schmidt-Weigand, Kohnert, & Glowalla, 2010). As a research methodology to investigate the allocation of attention during learning tasks, eye tracking can be versatile and incisive (Lee, Ke, & Ryu, 2023). Eye tracking offers a unique approach to testing aspects of technology-enhanced learning hypotheses, particularly concerning attentional processing during learning and relative to individual differences.

Many researchers have explored the relationship between eye movements and visuospatial processing during the encoding, maintenance, and retrieval of information (Bhatia, Mohite, Spataro, Rossi-Arnaud, & Mishra, 2021; Heuer, Ohl, & Rolfs, 2020; McAteer, McGregor, & Smith, 2023; Olivers & Roelfsema, 2020; Van der Stigchel & Hollingworth, 2018). For example, it was found that long-lasting memory traces could be achieved when the fixation duration on objects is longer during the study phase (Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015; Damiano & Walther, 2019; Liu, Shen, Olsen, & Ryan, 2017; Meghanathan, van Leeuwen, & Nikolaev, 2015; Schwedes & Wentura, 2016). In addition, Pertzov, Avidan, and Zohary (2009) revealed that more fixations during encoding improved memory performance. These findings indicate that features of fixations, such as frequency and duration, are related to memory capacity (Hannula et al., 2010; Olejarczyk, Luke, & Henderson, 2014; Wynn, Shen, & Ryan, 2019). Furthermore, several studies used fixations to measure attention distribution (Godfroid, Boers, & Housen, 2013; Godfroid & Uggen, 2013; Rayner, 2009) and predict subsequent memory performance (Chaffin, Morris, & Seely, 2001; Vainio, Hyönä, & Pajunen, 2009). Studies by van Ede, Chekroud, and Nobre (2019) and Williams, Pouget, Boucher, and Woodman (2013) reported that maintenance and retrieval of object features from visuospatial working memory are associated with gaze fixation locations on previously occupied objects. Saccadic eye movements are also related to visuospatial working memory (Ohl & Rolfs, 2017; Timm & Papenmeier, 2023; Van der Stigchel & Hollingworth, 2018). For example, saccadic eye movements were necessary to enhance visual feature encoding (Bays & Husain, 2008; Hanning, Jonikaitis, Deubel, & Szinte, 2016).

Unlike fixations and saccades, fixation disparity is a salient but currently underexplored eye tracking variable that can also be associated with visuospatial cognition. During a fixation movement, the line of gaze from both eyes does not align perfectly with the target of interest. This phenomenon is known as fixation disparity (fixation instability or vergence error, Jaschinski et al., 1999; Otero-Millan, Macknik, & Martinez-Conde, 2014). For those with normal vision, fixation disparity is typically negligible because the degree of fixation disparity is smaller than the region in perceptual space (called Panum's areas; Howard & Rogers, 1995; Ogle, Martens, & Dyer, 1967; Scheiman & Steinman, 1994). However, large fixation disparity can cause eye strain and cognitive discomfort (Mallett, 1974; Pickwell, 1989; Sheedy & Saladin, 1980). Otero-Millan et al. (2014) reviewed recent studies on fixational eye movements from the binocular vision perspective. Most studies in this analysis gathered that increasing fixation disparity caused decreased visual perception. Zaroff, Knutelska, and Frumkes (2003) suggested that measurements of stereo sensitivity could be an index of fixation disparity. Specifically, these scholars found that the acuity in stereovision decreased with aging but did not differ across genders. Most prior experiments were designed to quantify fixation disparities depending on the vergence stimulus, which makes both eyes in opposite directions to create a single binocular vision.

Interestingly, the relationship between fixation disparity and internal mental representation is debatable. While Benedek, Stoiser, Walcher, and Körner (2017) found that internally directed cognition (mental imagery, creative thinking, and mind wandering) led to a reduced angle of eye vergence for goal-oriented tasks, Walcher, Körner, and Benedek (2017) observed no difference in the angle of eye vergence between a reading task (which require externally directed cognition process for visual information processing) and an idea generation task (which involve internally directed cognition process). Annerer-Walcher, Körner, and Benedek (2018) also reported that fixation disparity remained unchanged despite the observed perceptual decoupling from external visual events, as evidenced by other eye movement behaviors. However, more studies focus on fixation disparity as an indicator of internal versus external attentional shift (Annerer-Walcher, Körner, Beaty, & Benedek, 2020; Ceh et al., 2021; Nascimento, Alvarez-Peregrina, Martinez-Perez, & Sánchez-Tena, 2021; Puig, Puigcerver, Aznar-Casanova, & Supèr, 2013; Puig, Zapata, Aznar-Casanova, & Supèr, 2013). Thus, it is reasonable to investigate whether poor fixation disparity affects visuospatial task performance by increasing reaction times and decreasing accuracy and its potential relationship to cognitive processes. Our research extends these prior works by exploring the relationship between fixation disparity and performance in an authentic organic chemistry educational task requiring visuospatial reasoning.

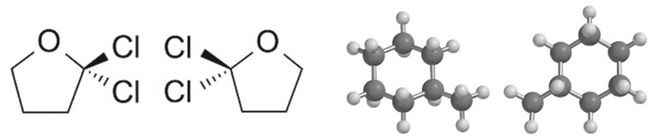

Fig. 1. A pair of dash–wedge (2D) models of molecules (left) and a pair of ball–stick (3D) models of molecules (right).

In organic chemistry, students often struggle with comparing molecules' visual and spatial characteristics to determine whether different molecule representations show the same or different molecules. Learners are also required to use various formats of representation to understand the structure of molecules. However, the unfortunate reality is that many educators do not prioritize the development of skills necessary for constructing, interpreting, and communicating representations of abstract concepts and relationships, also known as representational competence (Stieff et al., 2018).

This article introduces a novel method for quantifying fixation disparity from gaze positions in squared pixels. This method can be used to investigate the effects of fixation disparity in real-world visuospatial problem solving, such as learning in authentic and visuospatially rich environments (e.g., games and simulations, 3D and virtual reality spaces). We investigated fixation disparity during a common stereochemistry task in introductory organic chemistry courses. Participants compared juxtaposed pairs of two-dimensional (dash–wedge) and three-dimensional (ball–stick) molecular models and determined if they were superimposable or nonsuperimposable (i.e., chiral; Figure 1). A left pair of molecules is identical in 2D in Figure 1, while a pair of 3D molecules on the right differs. Given this educational context, our exploration was driven by the following research questions:

1. How does fixation disparity influence stereochemistry task performance when using two-dimensional (dash–wedge) and three-dimensional (ball–stick) molecular models?
2. Does the relationship between fixation disparity and performance on visuospatial cognitive tasks change over time?

REPRESENTING FIXATION DISPARITY USING COVARIANCE ERROR ELLIPSES

Measuring fixation disparity is not a new topic in the eye tracking research community. Fixation disparity can be quantified as both subjective and objective measurements. When viewing two nonius targets (short vertical/horizontal lines) presented separately to the left and right eye, a perceptional result is subjective fixation disparity. Those lines will be perceived as perfectly aligned, like a single line, with normal binocular vision. Jaschinski, Bröde, and Griefahn (1999) studied the relation between fixation disparity and the nonius (vernier) bias. To measure fixation disparity, they presented the nonius lines to participants who wore spectacle frames with polarizing filters so that the two lines were shown separately to the left and right eye. De Luca, Spinelli, Zoccolotti, and Zeri (2009) presented a computational method to calculate vergence angles and fixation disparity from gaze positions recorded by infrared eye trackers. Specifically, De Luca et al. (2009) suggested that the method requires an exact measure of viewers' interpupillary distance, viewing distance, and gaze positions. Objective fixation disparity is calculated by taking the vergence error of the visual axes when the target is viewed monocularly and when viewed binocularly. Jaschinski (2018) investigated the relationship between subjective and objective fixation disparities. Subjective fixation disparity refers to the perceptual test result when viewing two targets separately with only the left/right eye (Jaschinski, 2018). Objective fixation disparity measures the vergence error between monocular and binocular visual axes (Jaschinski, 2018). It was observed that if subjective fixation disparity is significant, objective fixation disparity is also large (Jaschinski, 2018). Previous studies on quantifying fixation disparity have predominantly focused on the sensory and motor systems within the human visual system. However, such studies often require clinical laboratory environments and physical measurements, which may not be practical in specific research contexts. Thus, it is necessary to propose an alternative approach to computing and comparing participants' fixation disparity.

Fixation disparity can also be quantified by measuring the distribution of gaze positions during fixation. Steinman (1965) introduced a method to measure fixation disparity by calculating an ellipse's area involving gaze positions during attempted fixation known for the bivariate contour elliptical area (BCEA). We decided to expand this method to a more practical setting, as it can be implemented using binocular pupil-CR (corneal reflection) video eye trackers, making it more ecologically valid. Our method extends Steinman (1965) by generating a participant's covariance error ellipse that contains 95% of the left- and right-eye difference of entire gaze positions during a visuospatial learning task. One fixation disparity feature sample can be defined as the absolute difference between the left- and right-eye gaze positions, $|G_L - G_R|$. Then, a covariance matrix is obtained using all fixation disparity feature values to draw a covariance error ellipse. Figure 2 shows a BCEA for the fixation disparity for two participants. The ellipse area indicates the participants' degree of fixation disparity, that is, the smaller the ellipse area, the less fixation disparity, and vice versa. It must be emphasized that the size of the covariance error ellipse per trial (including 95% of all gaze samples in one trial)

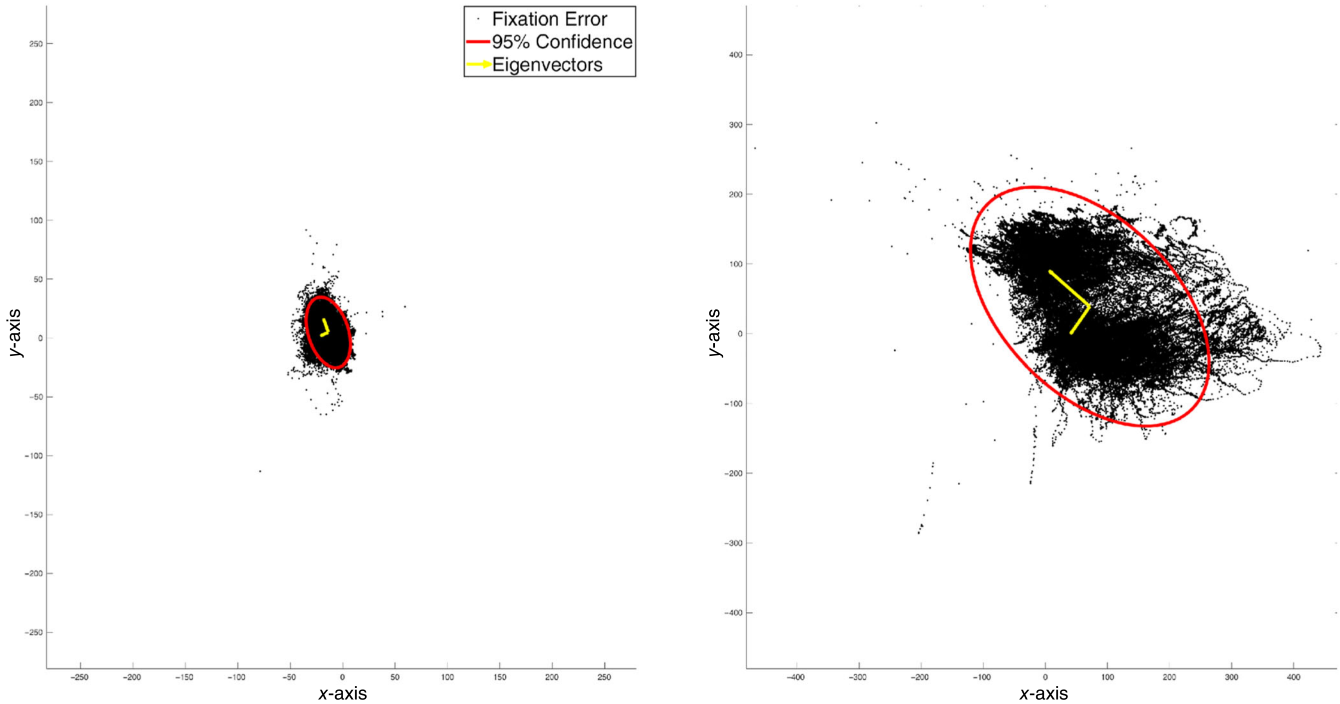

Fig. 2. The covariance error ellipse covering 95% of fixation disparity cases for two participants (left: small fixation disparity, right: large fixation disparity).

remained virtually unchanged during the experiment for the same subject. This observation indicates that fixation disparity and the system calibration error (less than 1° of visual angle by commercial eye tracking systems) remain constant. This observation also supports that the between-subject variability of our fixation disparity measurements is stable and reliable enough to explore individual differences that reflect Jaschinski et al.'s (1999) result when measuring subjective fixation disparity.

EMPIRICAL TESTING: STUDY 1

This study was designed to examine the influence of molecular model dimensionality (2D vs. 3D) on community college students' performance on a chirality task commonly used in introductory organic chemistry courses. We expected that Study 1 could reveal how fixation disparity influences stereochemistry task performance.

Participants

Fifty-five community college students in STEM degree programs were recruited for the study. All participants gave their informed consent before their inclusion in the study. The average age of the participants was 22.22 years old ($SD = 3.59$). There were 20 male students and 35 female students in the study sample. Only four participants had taken an organic chemistry course before, so most participants were novices relative to the experimental task. Unlike organic chemistry experts, who use diagrammatic reasoning heuristics to perform chirality determination, novices tend to rely on the visuospatial skills of mental rotation and pattern comparison when they perform chirality determination (Stieff, 2007). Thus, we expected the community college students to perform the experimental task differently based on their visuospatial skills.

Experimental Setup

The experiment employed a within-subjects design, and stimuli were custom-built in MATLAB Psychtoolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). There were two blocks of 80 trials. One of the blocks contained only 3D models of molecules (ball–stick), and the other block included only 2D models of molecules (dash–wedge) to compare. A pair of molecular models were presented to participants in each trial. Participants were asked to determine whether the two models represented the same or different molecules. Participants used the left and right keys on the keyboard to make their "Same" or "Different" selections. The left and right keys on the keyboard were labeled as "S" and "D" on top of their original labels to make them more apparent to the participants.

The duration of each trial was limited to 10 s. If a participant did not make a selection, the trial was marked as incorrect, and the participant was shown the next trial. Auditory feedback on the response was provided to participants as a high-pitch tone (correct) or low-pitch tone (incorrect). All stimuli were grayscale to control for the potential effect of color. A brief introductory video was provided for participants to explain the concept of chirality and introduce the task. The first 10 trials were reserved for a training session within each block. The presentation of each block was counterbalanced.

Data Collection

A video-based EyeLink 1,000 system recorded participants' binocular eye movements with a desktop-mounted setup at a 500 Hz sampling frequency. The EyeLink has an average accuracy of around 0.5°. A chin rest was used to increase the tracking stability. Participants had no prior experience with the eye tracking equipment. A nine-point calibration procedure was performed before each block. The experiment lasted about 30 min for each participant, including a break between blocks. Among the 55 participants, 45 participants' eye movement records were considered for analysis based on validity criteria (records with a data loss rate of less than 20% were considered). In addition, participant response times were recorded, and task accuracy was collected as a measure of chirality determination accuracy.

Behavioral Results

The maximum score for each dimensionality condition was 80 (10 practice trials plus 70 main trials). The average score for all participants was 59.07 ($SD = 9.24$) for the 2D condition and 55.53 ($SD = 8.81$) for the 3D condition. There was a statistically significant difference between the score means of the two-dimensionality conditions, $t(54) = 2.87$, $p < .01$. A paired sample t-test suggested that individual participants' scores on the 2D task were moderately correlated with their scores on the 3D task, $r(55) = 0.49$, $p < .001$. The average response time per trial for all participants was 3.53 s ($SD = 1.74$) for the 2D condition and 5.84 ($SD = 3.10$) for the 3D condition. Response time was significantly higher in the 3D condition than in the 2D condition, $t(54) = 8.31$, $p < .001$. Response times for 2D and 3D were strongly and positively correlated, $r(55) = 0.78$, $p < .001$. The score and response time were no statistically correlations for 2D ($r(55) = -0.19$, $p = .17$) and 3D ($r(55) = -0.12$, $p < .39$) conditions. These findings indicate that (1) participants' visuospatial cognition was more efficient in the 2D task condition (significantly higher scores and faster response times), (2) participants who did well in the 2D condition also did well in the 3D condition, and (3) there were no statistical correlations found between the score and response time for both conditions.

Fixation Disparity Results

The average area of error ellipses for the 2D condition was 13,900 square pixels ($SD = 472.08$) and 11,570 square pixels ($SD = 393.56$) for the 3D condition. The areas of error ellipses for both tasks were not normally distributed and it was right-skewed. A Wilcoxon signed-rank test showed no statistically significant difference in error ellipse size between tasks ($Z = -0.547$, $p = .585$), which implies that a participant's fixation disparities did not differ between tasks.

To investigate the relationship between task performance and fixation disparity, we proposed the task performance index using a ratio of accuracy/speed (A/S) calculated as $S_i / R_i * N$. Here, S_i is the score, R_i is the average response time, and N is the total number of trials for the participant i. Our A/S index is similar to the most often suggested measure called inverse efficiency score (Townsend & Ashby, 1985). This composite measure is commonly used to compare groups when controlling speed–accuracy trade-offs (Akhtar & Enns, 1989; Simon et al., 2008). Our A/S index was expected to be insensitive to speed–accuracy trade-offs because the behavioral results showed no significant negative correlation between score and response time (Liesefeld & Janczyk, 2018).

We divided participants into two groups based on the median A/S index because "median splits" are common practice in artificial categorization among clinical psychologists and other related researchers (DeCoster, Gallucci, & Iselin, 2011). Participants with lower than median A/S index (value of 0.22) were assigned to the "low" performing group and those with higher indices to the "high" performing group.

Twenty-two participants were in a high-performing group, and 23 were in a low-performing group. The mean AS index for the lower performing group was 0.15 ($SD = 0.04$) for the 2D condition and 0.10 ($SD = 0.03$) for the 3D condition. The mean AS index for the higher performing group was 0.36 ($SD = 0.14$) for the 2D condition and 0.20 ($SD = 0.08$) for the 3D condition (Figure 3).

A Mann–Whitney U test was performed to explore differences in the fixation disparity of low versus high performers (Figure 3) for both conditions. The fixation disparity for low performers ($M_{2DL} = 16.60e3$, $SD_{2DL} = 23.94e3$) was significantly greater than for high performers ($M_{2DH} = 10.70e3$, $SD_{2DH} = 12.27e3$) for the 2D condition, $U = 332{,}854$, $p < .001$. Fixation disparity for the low-performing group was also significantly higher for the 3D condition, $U = 258{,}871$, $p < .001$ ($M_{3DL} = 11.96e3$, $SD_{3DL} = 16.30e3$, $M_{3DH} = 11.11e3$, $SD_{3DH} = 14.84e3$). These findings indicate that fixation disparity could play a significant role because the fixation disparity was higher when the performance index was lower. Also, it is necessary to conduct a moderation analysis to explore whether fixation disparity affects task performance.

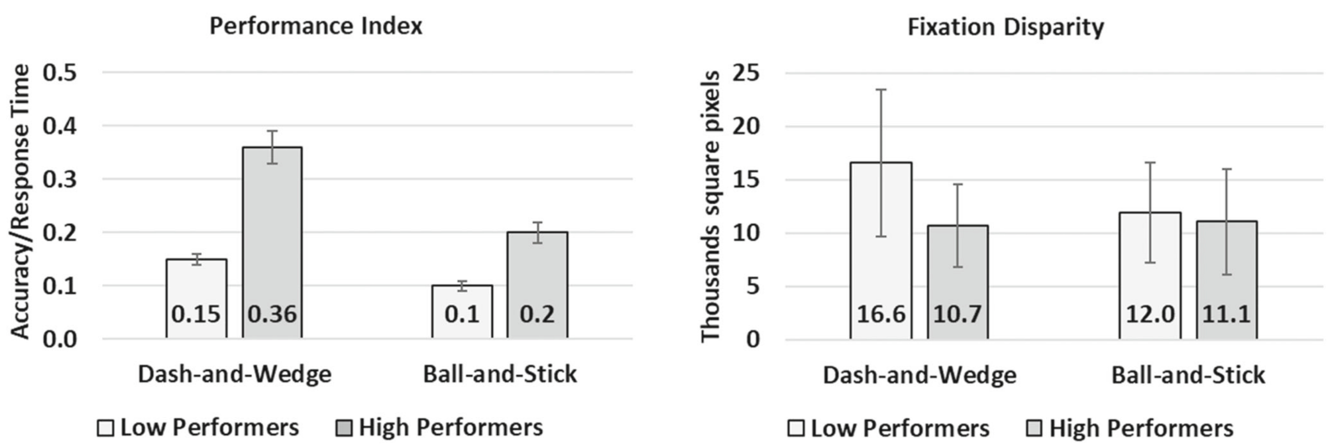

Fig. 3. Performance and fixation disparity comparisons for low and high performers in the 2D (dash-and-wedge) and 3D (ball-and-stick) tasks. Error bars indicate standard error.

EMPIRICAL TESTING: STUDY 2

The second study was designed to explore how sustained playing of the tile-matching game Mahjong, a popular pattern comparison game, can affect undergraduate students' visual attention patterns and performance on the chirality determination task, the same chirality task used in Study 1. Utilizing Mahjong as a cognitive leisure activity has shown a positive intervention effect on multiple cognitive functions, including attention, short-term memory, and working memory (Cheng, Chow, Song, Edwin, & Lam, 2014; Iizuka et al., 2018; Zhang et al., 2020). Furthermore, playing Mahjong may delay the progression of dementia and improve episodic memory (Cacciaglia et al., 2018; Cheng et al., 2014). However, previous studies did not examine participants' eye movement behaviors for any correlation with improved cognitive function. A crossword puzzle was chosen as a conventional treatment for Study 2 because an intervention effect on crossword puzzles was on phonic verbal fluency (Murphy, O'Sullivan, & Kelleher, 2014), which we believe is a non-related skill to the performance of the chirality task.

Participants and Experimental Setup

Participants had to meet the following inclusion criteria: (1) new to organic chemistry, (2) no history of neurological disorders and brain trauma, (3) between 18 and 30 years old, and (4) continued access to a smartphone. All participants gave their informed consent before their inclusion in the study. The average age of the participants was 20.37 years old (*SD* = 2.28). In this between-subjects design study, a sample of 59 students (32 female and 27 male) new to organic chemistry and the game of Mahjong were recruited from a large public university in the southeastern United States. Through random assignment, 18 females and 14 males were assigned to the Mahjong group (treatment), and 14 females and 13 males were assigned to the crossword group (control). The groups were deemed homogeneous relative to visuospatial working memory capacity and knowledge of organic chemistry because they (1) did not differ on their symmetry span task (Unsworth & Engle, 2008) scores and (2) had not taken an organic chemistry course prior to the study.

Participants completed the chirality task (presented in 2D and 3D) twice, once in Phase I and once in Phase II. During Phase I, after completing the chirality task in-lab, the treatment group was asked to install the free Mahjong app on their smartphones. The control group was asked to install the free Daily WorldCross app. Participants practiced playing one game of Mahjong (treatment) and completed one crossword puzzle (control). They were then instructed to play the respective game for at least 20 min a day and bring their phones to their in-lab Phase II visit, which took place after 1 month. Not all participants were from STEM majors, which impacted their task performance in Phase II less. Treatment fidelity was addressed by accounting for time played within a month (reported by the apps) as a covariate. During Phase II, time spent within the game since the first session and the total number of badges earned were collected as game engagement metrics. Participants were also required to complete the symmetry span test at each phase to measure visuospatial working memory capacity (Kane et al., 2004).

Behavioral Results

The total time spent within the game played was 389.67 min (*SD* = 205.66) for the treatment group and 369.72 min (*SD* = 213.51) for the control group. The total number of badges earned was 264.94 (*SD* = 186.80) for the treatment group and 93.93 (*SD* = 84.18) for the control group. Total time spent within the game was not statistically significant between groups. In Phase I, the symmetry span task score was 32 ($SD = 7.70$) for the treatment group and 31.68 ($SD = 5.63$) for the control group. In Phase II, the symmetry span task score was 35 ($SD = 5.40$) for the treatment group and 33.21 ($SD = 6.28$) for the control group. Analysis of covariance was performed with groups as a factor variable, total time spent within the game and the total number of badges earned as covariates, and the symmetry task performance as response variable. Analysis of covariance found the main effect of the treatment on participant performance on the symmetry span task, $F(1, 58) = 4.41$, $p = .04$.

Similar to Study 1, we compared the performance of both treatment and control groups by the task performance index using a ratio of A/S. A Wilcoxon signed-rank test showed that both groups' performance improved in both the 2D condition ($Z = 6.597$, $p < .001$) and the 3D condition ($Z = 6.393$, $p < .001$) after playing either smartphone game. For the 2D condition, the performance index of the treatment group for Phase I was 0.23 ($SD = 0.10$) and 0.34 ($SD = 0.13$) for Phase II. For the 3D condition, the performance index of the treatment group for Phase I was 0.13 ($SD = 0.05$). It was 0.22 ($SD = 0.09$) for Phase II. For the 2D condition, the performance index of the control group for Phase I was 0.24 ($SD = 0.09$) and 0.37 ($SD = 0.13$) for Phase II. For the 3D condition, the performance index of the control group for Phase I was 0.15 ($SD = 0.06$). It was 0.21 ($SD = 0.10$) for Phase II. Overall, both groups demonstrated statistically significant improvements in performance for the 2D and 3D task conditions in Phase II (posttest) compared to Phase I (pretest). The test results indicate that (1) participants responded quicker in Phase II compared to Phase I during the 2D task and the 3D task and (2) their accuracy scores were greater in Phase II compared to Phase I. The performance increase in Phase II is consistent with the results of Sowden, Davies, Rose, and Kaye (1996). Their results indicate that practicing stereoscopic tasks can significantly increase stereoscopic acuity. Our results imply that participants across both experimental groups demonstrated better performance on the 2D condition than on the 3D condition, confirming the findings of Study 1, which used the same chirality task with a different group of chemistry novices.

Fixation Disparity Results

A Mann–Whitney U test was performed to examine differences in the experimental groups' fixation disparity relative to test performance in Phases I and II. There were no statistically significant differences between groups for Phase I ($U = 481$, $p = .63$) and Phase II ($U = 469$, $p = .76$) in the 2D condition. There were no statistically significant differences between groups for Phase I ($U = 432$, $p = .82$) and Phase II ($U = 489$, $p = .55$) in the 3D condition. The fixation disparity for the control group ($M = 19.10e3$, $SD = 15.02e3$) in Phase I was slightly greater compared to Phase II ($M = 16.11e3$, $SD = 12.76e3$) for the 2D condition. The fixation disparity for the control group ($M = 18.68e3$, $SD = 18.20e3$) in Phase I was slightly smaller compared to Phase II ($M = 20.66e3$, $SD = 27.67e3$) for the 3D condition. The fixation disparity for the treatment group ($M = 22.52e3$, $SD = 26.45e3$) in Phase I was slightly greater compared to Phase II ($M = 17.37e3$, $SD = 20.23e3$) for the 2D condition. The fixation disparity for the treatment group ($M = 24.39e3$, $SD = 36.08e3$) in Phase I was also slightly greater compared to Phase II ($M = 15.35e3$, $SD = 14.42e3$) for the 3D condition. However, performance enhancements and fixation disparity values decreased between phases for the treatment (Mahjong) condition (Figures 4 and 5). A Wilcoxon signed-rank test showed a statistically significant performance improvement between the control group's phases in 2D ($Z = 4.55$, $p < .001$) and 3D ($Z = 4.42$, $p < .001$) conditions. A Wilcoxon signed-rank test also showed a statistically significant performance improvement between the treatment group's phases in 2D ($Z = 4.82$, $p < .001$) and 3D ($Z = 4.64$, $p < .001$) conditions. It should be emphasized that the performance increase of the treatment group in the 3D condition was significantly larger (74.43% increase) than other performance increase cases (48.11% increase).

There was no statistically significant correlation between the fixation disparity and total time spent in the game for the control group in the 2D condition ($r(28) = -0.08$, $p = .69$) and 3D condition ($r(28) = 0.19$, $p = .34$) for Phase I. There was no statistically significant correlation between the fixation disparity and total time spent in the game for the control group in the 2D condition ($r(28) = -0.01$, $p = .98$) and 3D condition ($r(28) = 0.02$, $p = .92$) for Phase II. There was no statistically significant correlation between the fixation disparity and total time spent in the game for the treatment group in the 2D condition ($r(32) = 0.16$, $p = .38$) and 3D condition ($r(32) = 0.13$, $p = .49$) for Phase I. There was no statistically significant correlation between the fixation disparity and total time spent in the game for the treatment group in the 2D condition ($r(32) = -0.01$, $p = .94$) and 3D condition ($r(32) = -0.003$, $p = .99$) for Phase II.

A multiple regression was run to predict test performance from fixation disparities for both phases. For Phase I of the control group, these variables did not statistically significantly predict the test performance in 2D condition, $F(2, 25) = 0.48$, $p = .63$, $R^2 = -0.04$. For the 3D condition, variables did not statistically significantly predict the test performance, $F(2, 25) = 2.33$, $p = .12$, $R^2 = 0.09$. For Phase I of the treatment group, these variables did not statistically significantly predict the test performance in 2D condition, $F(2, 29) = 1.08$, $p = .35$, $R^2 = 0.005$. For the 3D condition, variables did not statistically significantly predict the test performance, $F(2, 29) = 0.60$, $p = .56$, $R^2 = -0.03$.

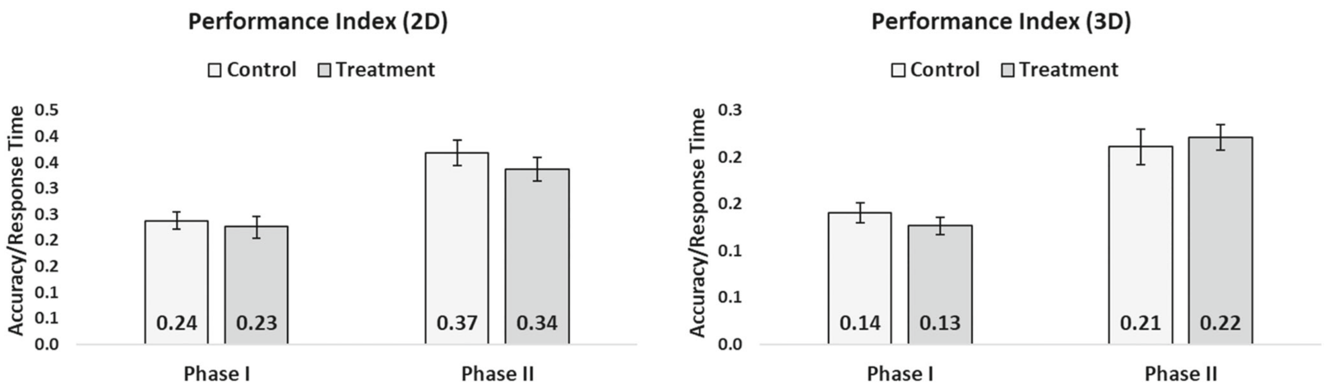

Fig. 4. Comparison of task performance between groups and phases. The treatment group played the Mahjong game (a pattern comparison game) for about 20 min each day. There was a 1-month gap between the phases. Error bars indicate standard error.

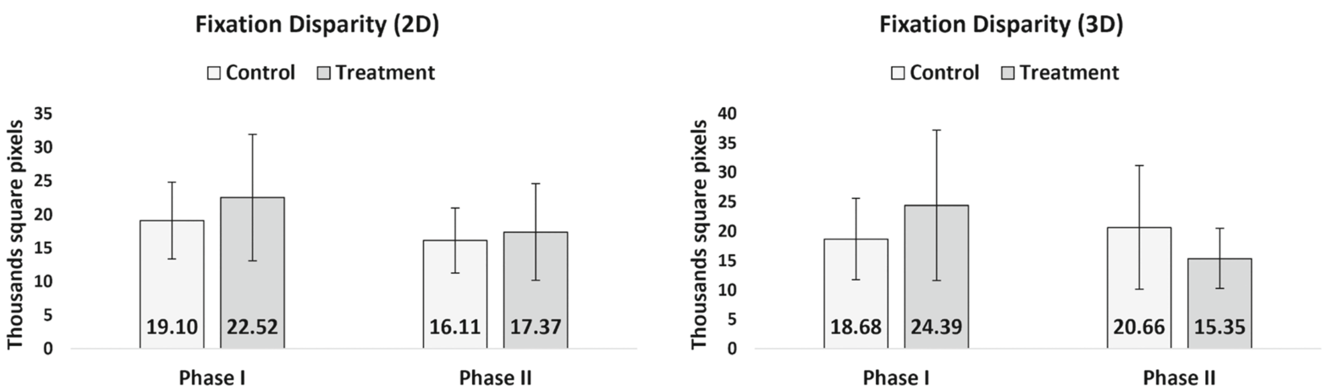

Fig. 5. Comparison of fixation disparity between groups and phases. Error bars indicate standard error.

For Phase II of the control group, these variables did not statistically significantly predict the test performance in 2D condition, $F(2, 25) = 1.08$, $p = .36$, $R^2 = 0.01$. For the 3D condition, variables did not statistically significantly predict the test performance, $F(2, 25) = 1.41$, $p = .26$, $R^2 = 0.03$. The total time spent in the game for the control group did not statistically significantly predict the test performance in the 2D condition ($F(1, 26) = 0.29$, $p = .59$, $R^2 = -0.03$) and in the 3D condition ($F(1, 26) = 0.10$, $p = .75$, $R^2 = -0.03$). For Phase II of the treatment group, these variables did not statistically significantly predict the test performance in 2D condition, $F(2, 29) = 0.03$, $p = .98$, $R^2 = 0.002$. For the 3D condition, variables did not statistically significantly predict the test performance, $F(2, 29) = 0.21$, $p = .82$, $R^2 = -0.05$. The total time spent in the game for the treatment group statistically significantly predicted the test performance in the 2D condition ($F(1, 30) = 14.60$, $p < .001$, $R^2 = 0.31$) not for the 3D condition ($F(1, 30) = 1.23$, $p = .28$, $R^2 = 0.01$).

To explore this further, we performed a nonparametric analysis of longitudinal data in factorial experiments suggested by Noguchi, Gel, Brunner, and Konietschke (2012), which is common in medical research, to explore the interaction between groups and phases. Noguchi et al. (2012) indicated that the method implements a broad range of rank-based nonparametric methods for analyzing longitudinal data in factorial experiments. Their approach assesses the treatment and time effects and their interactions via the analysis of variance-type statistics (ATS), accurately controlling the Type I error rate even for small sample sizes, and the classical Wald-type statistic (WTS). There were no statistically significant differences between groups ($p = .34$) and between groups and phases ($p = .56$) in terms of the performance index of the 2D condition. However, phases had a significant main effect on the performance index for the 2D condition based on both WTS and ATS with $p < .001$. There was also a significant interaction between groups and phases of the performance index for the 3D condition, $p < .05$ on both WTS and ATS. Finally, phase significantly affected the fixation disparity for the 2D condition, with $p < .05$ on both WTS and ATS (Table 1).

The results show that task performance in both 2D and 3D conditions increased significantly from Phase I to Phase II for both treatment and control groups. The treatment group showed more significant performance improvement (169% increase) than the control group (150% increase) for the 3D task. In addition, for the 2D task, fixation disparity

Table 1
The Summary of Tests of Between-Subjects Effects

	Effects on the 2D task performance		*Effects on the 3D task performance*		*Effects on the fixation disparity in 3D task*	
	Statistic	p	*Statistic*	p	*Statistic*	p
Group	0.92	.34	0.05	.83	0.13	.72
Phase	108.06	<.001*	102.38	<.001*	4.37	.04*
Group*Phase	0.33	.56	3.95	.047*	0.14	.71

The significance is reported for the following levels: *$p < .05$.

significantly decreased in Phase II for both groups. There was a more significant decrease in fixation disparity in the treatment group (22.88% decrease) than in the control group (15.68% decrease) for the 2D task condition, but not statistically significant.

DISCUSSION

We have started this article by discussing how numerous educational domains require students to tap into their visuospatial skills and how there is increasing evidence for the connection between visuospatial processing and domain-general problem-solving skills. While research is ongoing, we examined visuospatial processing in a novel and authentic organic chemistry learning task here. We have (1) proposed a novel approach to highlighting a specific eye tracking variable, fixation disparity, and (2) demonstrated how this eye tracking measure may relate to performance in an authentic, visuospatially rich learning task in two different samples of postsecondary students. Our method for measuring fixation disparity may be beneficial in many research contexts. Compared to most studies in social and behavioral sciences that use convenience samples of undergraduate students at their universities, in Study 1, we worked specifically with community college students, which improves the population validity of our approach (as 31% of the U.S. postsecondary student population consists of community college students; Irwin et al., 2022).

Study 1 revealed that participants' performance was better in the 2D condition than in the 3D condition, similar to Neubauer, Bergner, and Schatz's (2010) study on mental rotation performance with 2D or 3D objects. Higher performance in the 2D condition does not imply that 2D representations should always be preferred for stereochemistry problem-solving. Given that the 3D molecules in our experiment did not contain color or depth information or allow for virtual or physical manipulation, we cannot yet conclude that one representation is superior in all contexts. There is some evidence of the effectiveness of 3D stereoscopic models in education (Brown, Hamilton, & Denison, 2012; Cui, Wilson, Rockhold, Lehman, & Lynch, 2017; Nicholson, Chalk, Funnell, & Daniel, 2006; Qayumi et al., 2004). Valsecchi and Gegenfurtner (2012) even demonstrated that scene-viewing training using stereoscopic images contributes to retaining content in long-term memory, but only when participants took an extremely long time to comprehend the scene. It may be evident that, for our empirical studies, the performance of a task in a 2D molecule model is expected to be superior to one in 3D because the dash–wedge model is more simplified in this degree of detail and focuses more on only essential information for the task. On the other hand, the 3D model provides rich details that may unnecessarily increase cognitive load. Further studies may be required to compare the learning performance of different molecule models in 2D representation, 3D representation shown in 2D space, and 3D representation in virtual reality space, similar to the work of Rau, Michaelis, and Fay (2015) investigating the connection between multiple graphical representations in chemistry education.

Importantly, examining groups by performance in Study 1 also showed the possibility of the critical role of fixation disparity in visuospatial cognition—individuals with higher fixation disparity tended to perform worse on the authentic visuospatially rich learning task than their counterparts. Further research is needed to explore the relationship between fixation disparity and the types of cognitive processes required based on task conditions, as it differs between internal (mental rotation) and external (pattern matching) processes (Annerer-Walcher et al., 2020). Less fixation disparity of high performers could be aligned to the related studies that experts direct their attention more efficiently to relevant components of visualizations than novices (Hegarty, Canham, & Fabrikant, 2010; Jarodzka et al., 2010). Regarding implications for practice, the results from Study 1 suggest that chemistry educators could adapt and personalize learning experiences for students based on individual differences in fixation disparity measured from chirality learning tasks, similar to the role of saccade latency in diagnosing neurodegenerative diseases.

Study 2 revealed the relationship between participants' performance and fixation disparity of the two study phases. Specifically, participants' performance improved over time, and their fixation disparity measures decreased over time.

Our Study 2 was among the limited number of empirical investigations into how representational competence (a set of skills that involves understanding and using various forms of representations; Kozma & Russell, 2005) could be enhanced over time. Stieff, Hegarty, and Deslongchamps (2011) and Hinze et al. (2013) highlighted the importance of such empirical studies in STEM education. The results of Study 2 suggest that playing pattern-matching games, such as Mahjong, may influence learners' stereochemistry problem-solving skills. Some may argue that it may be less convincing because Mahjong may require only a direct pattern-matching skill, not a mental rotation skill. However, several studies mentioned that spatial skill training is not limited to the task (Feng, Spence, & Pratt, 2007; Kozhevnikov & Thornton, 2006). For instance, participants' practice gains from distinct spatial tasks were transferred to novel stimuli for the practiced and non-practiced tasks (Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). Several related studies also suggested that playing a suitable game might increase someone's visuospatial skill and take the role of perceptual training (stereo training) in recovering stereopsis or amblyopia (Ding & Levi, 2011; Levi & Li, 2009). A systematic review of intervention studies conducted by Iizuka et al. (2018) found that cognitive leisure activities could improve cognitive functions across domains, including working memory. This may suggest that participating in cognitive leisure activities that require mental rotation may lead to an even more significant improvement in performance, which could explain why the control group in Study 2 also showed a significant improvement in task performance. To address the limitations of the vague distinction between the control and treatment groups, we may need further study to introduce another group that does not participate in any cognitive activity. This group would be compared with our treatment and control groups to confirm the results.

Another finding from Study 2 was a significant performance increase and fixation disparity decrease in the treatment group (Mahjong) compared to the control group (crossword), especially for the 2D condition. This result was expected because the performance index analysis in Study 1 indicated that most participants performed poorly in the 3D condition compared to the 2D condition. We also expected that the treatment group showed more decreased fixation disparity than the control group because the 3D condition may require more mental rotation skills, and the 2D condition may require more pattern-matching skills needed to play Mahjong. It is important to conduct further investigations to understand how playing Mahjong improves task performance. The current study did not explore the relationship between the performance enhancement in Mahjong and task performance or individual differences in eye movement behavior. Therefore, additional research is needed to clarify this relationship. Instead, the result of Study 2 suggested that playing cognitive games could improve the measure of fixation disparities related to learning performance.

The findings from these two studies reveal a significant negative relationship between fixation disparity and visuospatial task performance in an authentic stereochemistry task, which is an important contribution to our understanding of the effects vision plays in visuospatially rich tasks. Our measure of fixation disparity is reasonably comparable to previous studies. Objective fixation disparity averaged about 4 min arc (about 28,055 squared pixels), and subjective fixation disparity averaged about 2 min arc (Erkelens, Steinman, & Collewijn, 1989; Jaschinski, 2017). Our studies also suggest that employing fixation disparity as a possible biomarker of visuospatial performance in educational research may be helpful. The research of Huang, Li, Ngai, Leong, and Bulling (2019) supports our idea by assuming that aberrant attentional shifting may hinder the conceptual learning process. Huang et al. (2019) also introduced the internal cognitive process detection based on observation of eye vergence behavior, which supports our emphasis on fixation disparity along with many other studies (Annerer-Walcher et al., 2020; Ceh et al., 2021; Nascimento et al., 2021; Puig, Puigcerver, et al., 2013; Puig, Zapata, et al., 2013). Our study findings strongly support that fixation disparity measured using our method is closely linked to visuospatial skills.

While both studies reveal the value in exploring fixation disparity in further educational contexts, we acknowledge that these investigations were not without limitations. One potentially physical limiting factor was using a table-mounted eye tracker to collect eye movement data. While necessary in the current experimental design, future studies should endeavor to replicate these findings with low-cost webcam-based eye trackers with no chin rest to allow for more naturalistic head and eye movements. As mentioned, Study 1 was limited by the decision to control for color information (typically present in 3D organic chemistry models) to allow for a more fair comparison with the 2D molecules. Additionally, Study 2 required a smartphone application to play over a month, which allowed us to investigate visuospatial processing changes following a relatively low time commitment and generally engaging intervention. However, it is possible that a more directed visuospatial intervention would have a different impact on fixation disparity and visuospatial performance. Thus, we acknowledge that our methodology should be probed more in further scientific contexts, with different smartphone games, and with organic chemistry experts to evaluate the method's utility in various learning environments.

CONCLUSION

Overall, we highlight the utility of examining underexplored eye tracking variables during learning. Specifically, we evaluated fixation disparity in the context of a visuospatial learning task common in organic chemistry courses—molecular chirality determination. Our two studies investigated a novel analysis of this eye tracking variable during the visuospatial task. They revealed individual differences between task performance and fixation disparity that changed over time. These insights provide crucial implications for using fixation disparity as a possible biomarker of visuospatial performance and further investigating fixation disparity in various educational contexts.

There is considerable support from various researchers regarding the significance of eye tracking variables in chemistry education. According to Rodemer, Eckhard, Graulich, and Bernholt (2020), noticeable differences exist in the eye tracking measures of beginner and advanced undergraduate chemistry students when performing case comparison tasks, such as fixation/transition ratio. Tóthová, Rusek, and Chytrý (2021) also explained that the fixation duration on areas of interest could provide insights into how students solve problems related to the periodic table. Recently, Langner, Graulich, and Nied (2022) suggested that eye tracking could be a helpful teaching tool for preservice education while creating multimedia learning material for chemical education. However, many previous studies overlooked the significance of fixation disparity in their data analysis. We wish to underscore the potential importance of our fixation disparity measure within the eye tracking research community and its application in chemistry education.

Acknowledgments—This material is based upon work supported by the NSF Science of Learning and Augmented Intelligence under Grant number 1540888.

CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest.

REFERENCES

Akhtar, N., & Enns, J. T. (1989). Relations between convert orienting and filtering in the development of visual attention. *Journal of Experimental Child Psychology*, *48*(2), 315–334.

Allen, K., Higgins, S., & Adams, J. (2019). The relationship between visuospatial working memory and mathematical performance in school-aged children: A systematic review. *Educational Psychology Review*, *31*(3), 509–531.

Alvarez-Vargas, D., Abad, C., & Pruden, S. M. (2020). Spatial anxiety mediates the sex difference in adult mental rotation test performance. *Cognitive Research: Principles and Implications*, *5*(1), 1–17.

Annerer-Walcher, S., Körner, C., Beaty, R. E., & Benedek, M. (2020). Eye behavior predicts susceptibility to visual distraction during internally directed cognition. *Attention, Perception, & Psychophysics*, *82*, 3432–3444.

Annerer-Walcher, S., Körner, C., & Benedek, M. (2018). Eye behavior does not adapt to expected visual distraction during internally directed cognition. *PLoS One*, *13*(9), e0204963.

Atit, K., Power, J. R., Pigott, T., Lee, J., Geer, E. A., Uttal, D. H., … Sorby, S. A. (2021). Examining the relations between spatial skills and mathematical performance: A meta-analysis. *Psychonomic Bulletin & Review*, *29*, 1–22.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Benedek, M., Stoiser, R., Walcher, S., & Körner, C. (2017). Eye behavior associated with internally versus externally directed cognition. *Frontiers in Psychology*, *8*, 1092.

Bhatia, D., Mohite, V., Spataro, P., Rossi-Arnaud, C., & Mishra, R. K. (2021). Effects of pointing movements on visuospatial working memory in a joint-action condition: Evidence from eye movements. *Memory & Cognition*, *50*, 1–17.

Boucheix, J. M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, *20*(2), 123–135.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Brown, P. M., Hamilton, N. M., & Denison, A. R. (2012). A novel 3D stereoscopic anatomy tutorial. *The Clinical Teacher*, *9*(1), 50–53.

Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*, 165–178.

Cacciaglia, R., Molinuevo, J. L., Sánchez-Benavides, G., Falcón, C., Gramunt, N., Brugulat-Serrat, A., … Huguet, J. (2018). Episodic memory and executive functions in cognitively healthy individuals display distinct neuroanatomical correlates which are differentially modulated by aging. *Human Brain Mapping*, *39*(11), 4565–4579.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166.

Castro-Alonso, J. C. (2019) *Overview of visuospatial processing for education in health and natural sciences*. (pp. 1–21). London: Springer eBooks. https://doi.org/10.1007/978-3-030-20969-8_1

Ceh, S. M., Annerer-Walcher, S., Koschutnig, K., Körner, C., Fink, A., & Benedek, M. (2021). Neurophysiological indicators of internal attention: An fMRI–eye-tracking coregistration study. *Cortex*, *143*, 29–46.

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 225.

Cheng, S. T., Chow, P. K., Song, Y. Q., Edwin, C. S., & Lam, J. H. (2014). Can leisure activities slow dementia progression in nursing home residents? A cluster-randomized controlled trial. *International Psychogeriatrics*, *26*(4), 637–643.

Cui, D., Wilson, T. D., Rockhold, R. W., Lehman, M. N., & Lynch, J. C. (2017). Evaluation of the effectiveness of 3D vascular stereoscopic models in anatomy instruction for first year medical students. *Anatomical Sciences Education*, *10*(1), 34–45.

Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, *184*, 119–129.

De Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, *20*(2), 111–122.

De Luca, M., Spinelli, D., Zoccolotti, P., & Zeri, F. (2009). Measuring fixation disparity with infrared eye-trackers. *Journal of Biomedical Optics*, *14*(1), 014013.

DeCoster, J., Gallucci, M., & Iselin, A. R. (2011). Best practices for using median splits, artificial categorization, and their continuous alternatives. *Journal of Experimental Psychopathology*, *2*(2), 197–209.

Ding, J., & Levi, D. M. (2011). Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), E733–E741.

van Ede, F., Chekroud, S. R., & Nobre, A. C. (2019). Human gaze tracks attentional focusing in memorized visual space. *Nature Human Behaviour*, *3*(5), 462–470.

Ehmann, P., Beavan, A., Spielmann, J., Ruf, L., Mayer, J., Rohrmann, S., … Englert, C. (2021). 360°-multiple object tracking in team sport athletes: Reliability and relationship to visuospatial cognitive functions. *Psychology of Sport and Exercise*, *55*, 101952.

Erkelens, C. J., Steinman, R. M., & Collewijn, H. (1989). Ocular vergence under natural conditions. II. Gaze shifts between real targets differing in distance and direction. *Proceedings of the Royal Society of London. B. Biological Sciences*, *236*(1285), 441–465.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850–855.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483–517.

Godfroid, A., & Uggen, M. S. (2013). Attention to irregular verbs by beginning learners of German: An eye-movement study. *Studies in Second Language Acquisition*, *35*(2), 291–322.

Halpern, D. F., & Collaer, M. L. (2005) *Sex differences in visuospatial abilities: More than meets the eye*. Cambridge: Cambridge University Press.

Hanning, N. M., Jonikaitis, D., Deubel, H., & Szinte, M. (2016). Oculomotor selection underlies feature retention in visual working memory. *Journal of Neurophysiology*, *115*(2), 1071–1076.

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*, 166.

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 37–53.

Heuer, A., Ohl, S., & Rolfs, M. (2020). Memory for action: A functional view of selection in visual working memory. *Visual Cognition*, *28*(5–8), 388–400.

Hinze, S. R., Williamson, V. M., Shultz, M. J., Williamson, K. C., Deslongchamps, G., & Rapp, D. N. (2013). When do spatial abilities support student comprehension of STEM visualizations? *Cognitive Processing*, *14*, 129–142.

Howard, I. P., & Rogers, B. J. (1995) *Binocular vision and stereopsis*. New York, NY: Oxford University.

Huang, M. X., Li, J., Ngai, G., Leong, H. V., & Bulling, A. (2019, October). Moment-to-moment detection of internal thought during video viewing from eye vergence behavior. *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2254–2262).

Iizuka, A., Suzuki, H., Ogawa, S., Kobayashi-Cuya, K. E., Kobayashi, M., Takebayashi, T., & Fujiwara, Y. (2018). Pilot randomized controlled trial of the G.O. game intervention on cognitive function. *American Journal of Alzheimer's Disease & Other Dementias*, *33*(3), 192–198.

Irwin, V., De La Rosa, J., Wang, K., Hein, S., Zhang, J., Burr, R., … Parker, S. (2022) *Report on the condition of education 2022 (NCES 2022-144). U.S. Department of Education*. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022144

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*(2), 146–154.

Jaschinski, W. (2017). Individual objective and subjective fixation disparity in near vision. *PLoS One*, *12*(1), e0170190.

Jaschinski, W. (2018). Individual objective versus subjective fixation disparity as a function of forced vergence. *PLoS One*, *13*(7), e0199958.

Jaschinski, W., Bröde, P., & Griefahn, B. (1999). Fixation disparity and nonius bias. *Vision Research*, *39*(3), 669–677.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *MPG PuRe (Max Planck Society)*, *36*(14), 1–16.

Knapp, A. (2011). Why schools don't value spatial reasoning. *Forbes*. Retrieved from https://www.forbes.com/sites/alexknapp/2011/12/27/why-dont-schools-value-spatial-reasoning/?sh=2f74c9b67b52

Kozhevnikov, M., & Thornton, R. (2006). Real-time data display, spatial visualization ability, and learning force and motion concepts. *Journal of Science Education and Technology*, *15*(1), 111–132.

Kozma, R., & Russell, J. (2005). Students becoming chemists: Developing representational competence. *Visualization in science education*, *1*, 121–146.

Langner, A., Graulich, N., & Nied, M. (2022). Eye-tracking as a promising tool in pre-service teacher education—A new approach to promote skills for digital multimedia design. *Journal of Chemical Education*, *8*(5), 65.

Lee, S., Ke, F., & Ryu, J. (2023). Engagement and effectiveness of symbolic and iconic learning support for math problem representation: An eye tracking study. *Interactive Learning Environments*, *31*(3), 1514–1531.

Levi, D. M., & Li, R. W. (2009). Perceptual learning as a potential treatment for amblyopia: A mini-review. *Vision Research*, *49*(21), 2535–2549.

Liesefeld, H. R., & Janczyk, M. (2018). Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods*, *51*(1), 40–60.
Liu, Z. X., Shen, K., Olsen, R. K., & Ryan, J. D. (2017). Visual sampling predicts hippocampal activity. *Journal of Neuroscience*, *37*(3), 599–609.
Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*(4), 344–351.
Mallett, R. F. J. (1974). Fixation disparity-its genesis and relation to asthenopia. *Ophthalmic Optician*, *14*, 1159–1168.
Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, *20*(2), 167–171.
McAteer, S. M., McGregor, A., & Smith, D. T. (2023). Oculomotor rehearsal in visuospatial working memory. *Attention, Perception, & Psychophysics*, *85*(1), 261–275.
Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience*, *8*, 1063.
Munion, A. K., Stefanucci, J. K., Rovira, E., Squire, P., & Hendricks, M. (2019). Gender differences in spatial navigation: Characterizing wayfinding behaviors. *Psychonomic Bulletin & Review*, *26*(6), 1933–1940.
Murphy, M., O'Sullivan, K., & Kelleher, K. G. (2014). Daily crosswords improve verbal fluency: A brief intervention study. *International Journal of Geriatric Psychiatry*, *29*(9), 915–919.
Nascimento, H., Alvarez-Peregrina, C., Martinez-Perez, C., & Sánchez-Tena, M. Á. (2021). Differences in visuospatial expertise between skeet shooting athletes and non-athletes. *International Journal of Environmental Research and Public Health*, *18*(15), 8147.
Neubauer, A. C., Bergner, S., & Schatz, M. (2010). Two-vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence*, *38*(5), 529–539.
Nicholson, D. T., Chalk, C., Funnell, W. R. J., & Daniel, S. J. (2006). Can virtual reality improve anatomy education? A randomized controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education*, *40*(11), 1081–1087.
Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, *50*(12), 1–23.
Ogle, K. N., Martens, T. G., & Dyer, J. A. (1967) *Oculomotor imbalance in binocular vision and fixation disparity*. Philadelphia, PA: Lea and Febiger.
Ohl, S., & Rolfs, M. (2017). Saccadic eye movements impose a natural bottleneck on visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 736.
Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes from eye movements. *Visual Cognition*, *22*(7), 975–995.
Olivers, C. N., & Roelfsema, P. R. (2020). Attention for action in visual working memory. *Cortex*, *131*, 179–194.
Otero-Millan, J., Macknik, S. L., & Martinez-Conde, S. (2014). Fixational eye movements and binocular vision. *Frontiers in Integrative Neuroscience*, *8*, 52.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.
Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10), 2.
Pickwell, D. (1989) *Binocular vision anomalies: Investigation and treatment*. London, England: Butterworths.
Puig, M. S., Puigcerver, L., Aznar-Casanova, J. A., & Supèr, H. (2013). Difference in visual processing assessed by eye vergence movements. *PLoS One*, *8*(9), e72041.
Puig, M. S., Zapata, L. P., Aznar-Casanova, J. A., & Supèr, H. (2013). A role of eye vergence in covert attention. *PLoS One*, *8*(1), e52955.
Qayumi, A. K., Kurihara, Y., Imai, M., Pachev, G., Seo, H., Hoshino, Y., … Kariya, Y. (2004). Comparison of computer-assisted instruction (CAI) versus traditional textbook methods for training in abdominal examination (Japanese experience). *Medical Education*, *38*(10), 1080–1088.
Rau, M. A., Michaelis, J. E., & Fay, N. (2015). Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers & Education*, *82*, 460–485.
Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.
Rodemer, M., Eckhard, J., Graulich, N., & Bernholt, S. (2020). Decoding case comparisons in organic chemistry: Eye-tracking students' visual behavior. *Journal of Chemical Education*, *97*(10), 3530–3539.
Scheiman, B. W. M., & Steinman, I. B. (1994) *Clinical management of binocular vision*. Philadelphia, PA: Lippincott.
Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and Instruction*, *20*(2), 100–110.
Schwedes, C., & Wentura, D. (2016). Through the eyes to memory: Fixation durations as an early indirect index of concealed knowledge. *Memory & Cognition*, *44*(8), 1244–1258.
Sheedy, I., & Saladin, I. I. (1980). Validity of diagnostic criteria and case analysis. *American Journal of Optommetry and Physiological Optics*, *57*(9), 618–631.
Simon, T. J., Takarae, Y., DeBoer, T., McDonald-McGinn, D. M., Zackai, E. H., & Ross, J. L. (2008). Overlapping numerical cognition impairments in children with chromosome 22q11. 2 deletion or Turner syndromes. *Neuropsychologia*, *46*(1), 82–94.
Sowden, P. T., Davies, I., Rose, D., & Kaye, M. (1996). Perceptual learning of stereoacuity. *Perception*, *25*(9), 1043–1052.
Steinman, R. M. (1965). Effect of target size, luminance, and color on monocular fixation. *Journal of the Optical Society of America*, *55*(9), 1158–1164.
Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*(2), 219–234.
Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying representational competence with multi-representational displays. *Cognition and Instruction*, *29*(1), 123–145.
Stieff, M., Origenes, A., DeSutter, D., Lira, M. E., Banevicius, L., Tabang, D. N., & Cabel, G. (2018). Operational constraints

on the mental rotation of STEM representations. *Journal of Educational Psychology*, *110*(8), 1160–1174.

Timm, J. D., & Papenmeier, F. (2023). Processing spatial configurations in visuospatial working memory is influenced by shifts of overt visual attention. *PLoS One*, *18*(2), e0281445.

Tóthová, M., Rusek, M., & Chytrý, V. (2021). Students' procedure when solving problem tasks based on the periodic table: An eye-tracking study. *Journal of Chemical Education*, *95*(3), 55.

Townsend, J. T., & Ashby, F. G. (1985). Stochastic modeling of elementary psychological processes. *American Journal of Psychology*, *98*(3), 480.

Unsworth, N., & Engle, R. W. (2008). Speed and accuracy of accessing information in working memory: An individual differences investigation of focus switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 616–630.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352.

Vainio, S., Hyönä, J., & Pajunen, A. (2009). Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading. *Experimental Psychology*, *56*(1), 66–74.

Valsecchi, M., & Gegenfurtner, K. R. (2012). On the contribution of binocular disparity to the long-term memory for natural scenes. *PLoS One*, *7*(11), e49947.

Van der Stigchel, S., & Hollingworth, A. (2018). Visuospatial working memory as a fundamental component of the eye movement system. *Current Directions in Psychological Science*, *27*(2), 136–143.

Walcher, S., Körner, C., & Benedek, M. (2017). Looking for ideas: Eye behavior during goal-directed internally focused cognition. *Consciousness and Cognition*, *53*, 165–175.

Williams, M., Pouget, P., Boucher, L., & Woodman, G. F. (2013). Visual–spatial attention aids the maintenance of object representations in visual working memory. *Memory & Cognition*, *41*, 698–715.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review*, *15*, 763–771.

Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic content. *Vision*, *3*(2), 21.

Zaroff, C. M., Knutelska, M., & Frumkes, T. E. (2003). Variation in stereoacuity: Normative description, fixation disparity, and the roles of aging and gender. *Investigative Ophthalmology & Visual Science*, *44*(2), 891–900.

Zhang, H., Peng, Y., Li, C., Lan, H., Xing, G., Chen, Z., & Zhang, B. (2020). Playing mahjong for 12 weeks improved executive function in elderly people with mild cognitive impairment: A study of implications for TBI-induced cognitive deficits. *Frontiers in Neurology*, *11*, 178.