Spatial Ability and Working Memory in STEM Learners

By

Nickita Gupta

Thesis Submitted in Partial Fulfillment of

the Requirements for the College of Letters and Science

Biopsychology Honors Program

In the Psychological and Brain Sciences Department,

Faculty of Applied Science

University of California, Santa Barbara

Abstract

It has been well documented that high spatial ability is a predictor of success in STEM; however, the cognitive mechanisms of spatial ability that build the foundation of learning and development of expertise have remained elusive. It has been theorized that experts are able to use their domain-specific knowledge to chunk semantically relevant information together in order to seemingly expand their visuospatial working memory. Novice learners of a discipline as well as those unfamiliar to the domain altogether are thought to rely on domain-general patterns and symmetry to compress the information. This study looks to better understand whether novice learners studying chemistry are building expert-level knowledge. Furthermore, we investigate whether spatial ability itself, apart from chemistry education, predicts accuracy in change detection paradigms that implement tasks more demanding of spatial working memory than previously studied. In Part 1, participants were presented with a ball-and-stick molecular structure stimulus that disappeared after a brief encoding period and then reappeared following a static mask as a target stimulus identical or different to the cue. The mismatch trials contained single elemental color changes or color swaps (switching two elemental positions within the molecule) each with rotations. In Part 2 and 3, participants completed Paper Folding and Cube Comparisons spatial ability measures, respectively. Our results indicate that overall, novices performed with higher accuracy in all trials of the change detection task and in both spatial measures than naive participants. Furthermore, we found that individuals with high spatial ability also performed with high accuracy. These two independent effects on accuracy suggest that chemistry education does play a significant role in determining spatial working memory change detection accuracy.

Key Words: visuospatial working memory; spatial ability, chemistry education; change detection

Introduction

The act of giving someone directions or retracing steps to remember where an item might be misplaced illustrates spatial thinking's role in producing everyday action. The application of spatial thinking, however, extends beyond daily tasks. An early book from Super and Bachrach (1957) titled “Scientific Careers and Vocational Development Theory” outlines how major STEM disciplines like biology, mathematics, and physics demand high spatial ability (Super & Bachrach, 1957, Shea, Lubinski & Benbow, 2001). These disciplines often require individuals to visualize properties of size, shape, movement, distance, and change over time and mentally transform the spatial representations created (Stieff et al., 2020). The integration and transformation of spatial properties make demands on visuospatial working memory. The practical applications of these visuospatial working memory object manipulations form the foundation of STEM disciplines. From the use of geographic information systems in mapping terrains, to the ability to rotate chemical molecules in space to understand and predict their pharmacological effects, spatial thinking remains a critical cognitive function in producing these scientific actions (Stieff et al., 2020). Much of the existing previous research explores how spatial ability can be a predictor of success in STEM. However, many questions regarding the cognitive mechanisms underpinning spatial thinking and ability in STEM remain unanswered (Uttal et al., 2013, Wai, Lubinski, & Benbow, 2009) .

The average visuospatial working memory capacity is three to four units of information (Luck & Vogel, 1997, Alvarez & Cavanaugh, 2004). Interestingly, as this information gets transformed as in STEM disciplines, our capacity for visuospatial working memory becomes more and more limited (Xu & Franconeri, 2015). In change detection paradigms involving a cross-shaped “molecule-like” stimulus made up of four different colors, Xu and Franconeri tested participants in their ability to mentally rotate the object by testing their accuracy in detecting a change in a stimulus following a rotation. For instance, in the change detection paradigm, one element of the original cross-shaped stimulus would change colors, and the entire stimulus would rotate. They found that our human visual system isn't optimal for maintaining multiple chunks and changes occurring simultaneously (Xu & Franconeri, 2015). Yet, the complexities of information transformations don't limit the level of expertise an individual can achieve within a discipline. So, how does expertise form and manifest itself in spatial ability task performance?

An understanding of the strategies experts utilize in non-STEM instances can give insight into similar strategies underlying cognitive mechanisms that are utilized within STEM disciplines. In the 1973 *Perception in Chess* study, Chase and Simon examined chess masters in their ability to recall complex configurations of chess pieces and compared it to novice chess players' performance. Results indicated that experts and novice players performed similarly in recalling nonsense formations, but experts repeatedly performed better when the information was semantically relevant and valid to the gameplay of chess (Chase & Simon, 1973). This cognitive strategy of grouping semantically relevant information together and representing it as a single unit of information became known as domain specific chunking, or expert chunking (Miller, 1956). Because of their familiarity with chess, masters could effectively chunk meaningful chess patterns together and therefore store more information in their limited visuospatial working memories. This strategy demonstrates the role of domain knowledge in drawing semantic connections that novices and naives unfamiliar to the game of chess may not be able to recruit (Chase & Simon, 1973).

However, it must be noted that expertise and the domain-specific chunking strategy may not be the only explanation for good performance in spatial working memory tasks. Stieff et al. (2020) addressed this possibility as they used a change detection paradigm with ball-and-stick molecular models to understand how chemistry novices (those who have taken at least introductory college-level organic chemistry) and naives (those unfamiliar to organic chemistry) detect specific types of changes. By testing novice and naive individuals, they were able to explore whether novice learners are starting to gather expertise strategies, or whether they perform similarly to naive individuals and thus recruit non-expert strategies in these recall tasks. Based on whether and what kind of expert-level strategies novices may be able to variably recruit over naive participants, we can observe the role that chemistry education plays in spatial ability performance tasks.

The change detection task in Stieff et. al (2020) used variations of ball-and-stick molecular structures that are often used to visually depict chemical molecules, with each colored ball representing a different atom and individual (or groups) of ball(s) connected to the central atom referred to as *ligands* in organic chemistry. The *ligands,* or atom(s) that attach to a central atom, were coded to be common *functional groups,* or common sets of atoms that are familiar to experts (expert chunks). In the change detection task, a ball-and-stick cue stimulus appeared briefly and was followed by a target stimulus that was either identical or a mismatch, each presented with 10 degree rotation from the cue. Mismatch trials featured stimuli with a single atom color change from the cue stimulus, and the color change took place either within a chemically salient functional group (expert chunk) or in the neighboring atoms, preserving the functional group. Identifying these chunk-changing and chunk-maintaining manipulations could serve as an index of expert-level encoding strategies being employed by learner populations. For example, if knowledge plays a role in encoding strategies, a common functional group like methyl (CH_3) could be represented and coded as a group of four atoms with two different colors (one atom colored for the carbon (C) and three atoms of one color for the hydrogens (H)). A chunk-changing mismatch would disrupt the color scheme of this common methyl functional group allowing those with expert-strategies to notice that the chemical composition of the methyl represented is no longer methyl. A chunk-maintaining mismatch would allow an individual with expert-strategies to quickly recognize the methyl group expert chunk first, and after confirming it is in fact still a methyl, distribute attention to analyzing other parts of the structure for changes. Stieff et. al found that overall Identical trials were easier to detect than Mismatch trials. Interestingly, when considering mismatch trials exclusively, results indicated that both novices and naives were more able to detect color changes within chemically relevant chunks (chunk changing manipulations), than chunk maintaining manipulations. Because both novices and naives performed with similar accuracy, this finding suggested that additional cognitive mechanisms other than domain-specific chunking could be at play.

One possibility is that general spatial ability rather than expertise contributes to higher accuracy in spatial tasks. However the accuracy of the color change detection task was not heavily correlated with performance on spatial ability tasks like Paper Folding and Cube Comparisons, suggesting that color changes maybe weren't the best measure of spatial manipulation. These findings prompt further exploration of other kinds of change detections that are more spatial in nature such as atoms (balls) switching places in the molecular model with additional rotation of the molecule.

Furthermore, higher performance on measures of working memory involving STEM stimuli may be independent of expertise and domain familiarity. There may be other methods to manage visuospatial representations in order to manipulate them in visuospatial working memory. The alternative domain general compression strategy states that individuals can detect repetition, symmetry, and spatial organization within visuospatial stimuli and use these properties to group and therefore enhance the amount of information in their visuospatial working memory capacity (Brady, Konkle, & Alvarez, 2009). Hence, Stieff et al's results of both novices and naives having similar sensitivity to color change detections, especially within expert chunks, supports initial evidence of this domain-general compression being a plausible strategy (Stieff et al., 2020).

These two competing strategies of domain general compression and domain-specific chunking prompt research testing learners of STEM and non-STEM disciplines in more spatially-oriented change detection tasks. A high accuracy performance in change detection tasks by learners of both groups could suggest that this domain-general cognitive strategy is being recruited, as neither group has developed true expertise. However, improved accuracy in novices could suggest that learners in STEM have begun accumulating the knowledge base and tools of expertise in their undergraduate education to utilize domain specific chunking. Ultimately, these change detections can provide insight and guidance upon the development of digital learning platforms for chemistry and can be a basis for pedagogical improvements in teaching strategies.

This current study's purpose is to accumulate more evidence on the role that chemistry education with subsequent learner expertise and spatial ability play in performing STEM stimuli change detection tasks. In doing so, it is essential to ensure that the change detection tasks created for this study are demanding of spatial working memory rather than visual working memory.

This study expands on the research of Stieff et al. (2020) to look at change detection through the lens of a more spatial working memory task. Specifically, it uses a change-detection task to examine element swaps (a spatial change) and compares these to element changes (a visual change) with both types of changes occurring with 10 and 120 degree rotation to ensure a greater degree of spatial manipulation. To support that element swaps are in fact more spatial than visual color changes, a correlation between the performance in the change detection task and traditional spatial ability tasks of Paper Folding and Cube Comparisons from the ETS' Kit of

Factor-Referenced Tests (Ekstrom, French, & Harman, 1976) will be conducted. Furthermore, effects of each type of mismatch trial on accuracy will help determine the relative difficulty of specific types of change detections. These findings can influence the future of STEM-specific spatial research and can have major pedagogical implications of integrating effective perceptual learning techniques when determining the design of online programs that truly deliver consistent STEM skill outcomes (Kellman, Massey, & Son, 2010, Shea, Lubinski, Benbow, 2001, Stieff & Uttal, 2015).

The Present Experiment

The objective of this study is to understand how well novice and naive students encode and transform chemistry ball-and-stick molecular model stimuli in change detection paradigms. Each participant will be presented with a cue encoding molecular structure stimulus with four unique ligands attached to the central atom, one of which will be a chunk. **See Figure 1**.

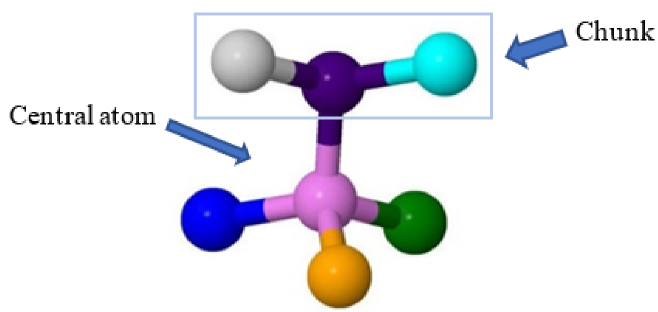


Figure 1. Example ball-and-stick molecular structure with a pink central atom connected to four ligands, one of which is a chunk

After initial presentation of the cue, participants will be shown a target stimulus and asked whether the target and cue are identical or a mismatch. The mismatch target stimuli will vary by the change in color of a single sphere (referred to as “color change” hereafter) or the switching of two spheres (or groups of spheres) on the model (referred to as “color swap” hereafter), with each mismatch being presented with a 10 or 120 degree rotation. Based on Xu and Franconeri's work, we predict that color swaps, especially when rotated, will be harder to detect than color changes because these mismatches are more demanding of spatial working memory.

If there exists no distinction between how naive and novice students perform, it may be due to an enhanced sensitivity in change detection around spatial groups, regardless of domain (Stieff et al., 2020). If a distinction is found, this may suggest that novice learners are developing perceptual knowledge towards familiarity, and taking potential steps toward a level of expertise.

To gauge baseline spatial ability, participants will also complete the standard cognitive spatial ability tasks of Paper Folding and Cube Comparisons (Ekstrom, French, & Harman, 1976). This will help determine which individuals have high spatial ability and understand if their ability translates across to STEM stimuli change detection paradigms. It is predicted that participants who are of high spatial ability (determined by Part 2) in both novice and naive groups will perform well in the various STEM stimuli change detection tasks that have been modified to feature more spatial transformations than previous research. Furthermore, we predict that due to the recruitment of spatial working memory in color swaps, spatial ability will be more highly correlated with color swaps than color changes.

Methods

Participants

41 undergraduate students who major in STEM disciplines and have taken at least one quarter of introductory organic chemistry (novices) and 41 undergraduate students with no organic chemistry background (naives), both from the University of California, Santa Barbara, participated in this study. Naive participants were recruited from the UCSB SONA subject pool

and received course credit for participation, while novices were recruited from outreach to introductory organic chemistry courses and received $10 Amazon Claim codes as compensation for participation.

Design

The design of the change detection paradigm was a 2 (presence of change: change vs. no change) x 2 (expertise: naive vs. novice) x 2 (type of manipulation: color change vs. color swap) x 2 (placement of change: chunk-maintained vs chunk-changed) x 2 (rotation: 10 degree vs 120 degree) mixed design.

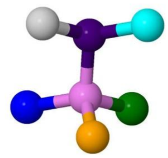
A linear mixed model with the above predictors, accuracy in change detection as the dependent variable, and participants as a random factor were fitted to the data. Additionally, several correlations between specific mismatch trials and spatial ability measures were calculated.

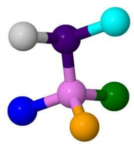
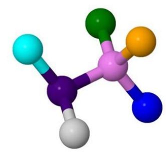
Materials. The study materials included a change detection paradigm using ball-and-stick molecular models rendered in JMol™, a chemistry modeling software, spatial ability measures such as the Paper Folding and Cube Comparison test, and a post-test survey on participant demographics and educational information.

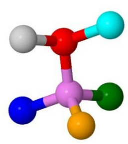
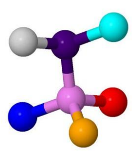
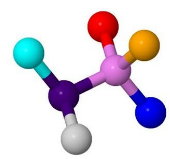
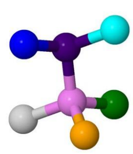
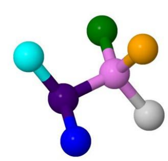
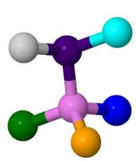
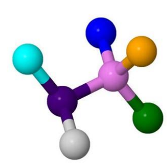
Molecular Representation Stimuli. The ball-and-stick molecular stimuli were created and rendered in the chemistry digital drawing program Jmol™ which uses shading and perspective to generate pseudo-3D images (Stieff et al., 2020). Each stimulus was composed of a central element that was connected to four other *ligands*, three of which were single spheres and the remaining one a chunked element (consisting of three spheres). There were a total of 6 elements (each sphere representing an atom) that varied in color and placement across the four ligands. The central element remained constant in color and placement in each molecular presentation. Standard color conventions of the JMol program were used to differentiate atomic identity.

Participants were sequentially presented with an encoding stimulus followed by a target stimulus. Eight (8) unique encoding stimuli (pseudo-3D molecules) were created. The target stimulus could either be a match, identical to the encoding stimulus, or a mismatch. Match targets were identical to the initial encoding stimuli. Mismatch targets were of two types: *color changes* or *color swaps* with half of each mismatch trial containing a chunk maintaining change while the other half containing a chunk-changing change. Specifically, in *color change mismatches* one single element (sphere) either in-chunk or as an independent ligand was altered in color from the encoding stimulus. Changes that disrupted the make-up of the encoding stimulus chunk were further classified as *chunk-changed color change mismatches*, while changes that maintained the makeup of the encoding stimulus chunk were classified as *chunk-maintained color change mismatches*. In *color swap mismatches*, two single elements (e.g., single atom-single atom or single atom-chunk) were switched in spatial placement. Swaps that disrupted the make-up of the encoding stimulus chunk were further classified as *chunk-changed color swap mismatches*, while changes that maintained the makeup of the encoding stimulus chunk were classified as *chunk-maintained color swap mismatches*. For all types of trials (matches and mismatches) half of the stimuli were rotated 10 degrees from the encoding stimulus to the target and half were rotated 120 degrees from the stimulus to the target, with rotations equally distributed clockwise and counterclockwise. **Figure 2.** below illustrates the different match and mismatch stimuli presented. Each of the eight (8) unique encoding stimuli and their variations were sequentially presented online within a specified bounds against a white background.

Figure 2. *Example of Various Target Stimuli for below encoding stimuli:*



Match Targets	10° rotation	120° rotation
Identical		

Mismatch Targets	10° rotation	120° rotation
Chunk-changing color change		
Chunk-maintaining color change		
Chunk-changing color swap		
Chunk-maintaining color swap		

Structure Change Detection Task. In the change detection task, participants were shown an encoding stimulus and then a sequential presentation of an identical target or a mismatch target, separated by a brief static screen delay. There were 16 trials per 8 unique encoding stimuli (eight match and eight mismatch presentations) for a total of 128 total trials.

Psychometric Measures of Spatial Ability. The Paper Folding and Cube Comparisons tests were administered following the change detection procedure.

In the Paper Folding Task (Ekstrom et al., 1976), participants were presented with a digital piece of paper that was folded in a series of steps. In the last step, a hole was punched on the folded sheet of paper. Participants were to select the correct option (out of 5) that illustrated the correct placement of the hole(s) once the paper was unfolded. Participants had a total of six minutes to complete twenty problems (3 minutes for each ten problems per page). The test was scored by taking the number of correct answers minus one fourth of all incorrect answers.

The Cube Comparisons Task (Ekstrom, et al., 1976) displayed two cubes with various letters or numbers on each face of the cube, with no two faces sharing the same number or letter. Participants were asked to determine if the two cubes were the same or different. There were a total of 21 items on each page, and participants were given three minutes to complete each of the two pages. The test was scored based on the number of correct answers minus the number of incorrect answers.

Participant Demographic Survey. We collected demographic and education information among participants with a post-test survey hosted and administered through Qualtrics™. Questions are listed in Appendix A.

Procedure

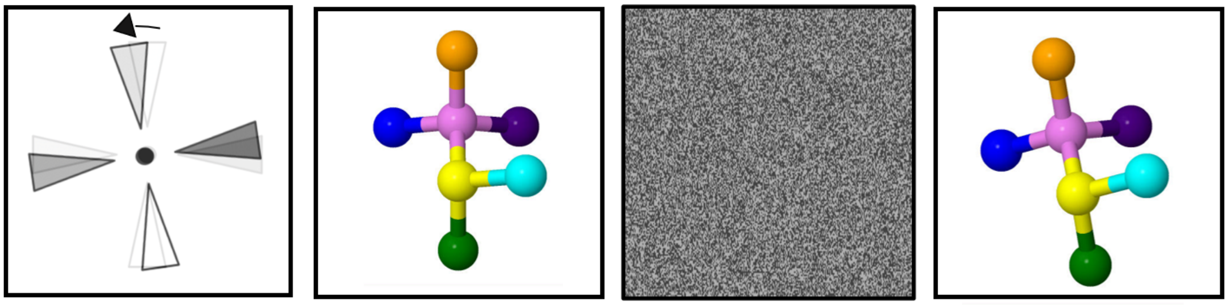
Each participant that signed up for the experiment was redirected to the online version of the experiment presented on PsychoPy. Once participants provided informed consent to continue, they were led through the instructions for the change detection procedure.

The instructions stated that participants would be presented with structures, but did not explicitly mention any keywords such as “chemical representations" or “molecular structures". They were told that these structures would be initially presented and then would disappear and be replaced with first a static screen (a mask) and then a new structure. Prior to each initial structure presentation, participants saw a rotating windmill that indicated how much the second structure was rotated compared to the first. Participants were instructed to respond by the keystrokes of “1” if they judged the two structures to be different or “9” if they judged the structures to be the same. Following these instructions, participants were led through six practice trials with embedded feedback that notified participants whether their answers were correct or incorrect or if they needed to respond faster.

Following the practice trials, participants were taken to the randomized 128 trials of the change detection procedure. After completing all the trials, participants advanced to Part 2 and completed the Paper Folding Task. Upon completion participants were advanced to Part 3 to complete the Cube Comparisons Task. After completing the two spatial ability measures, participants were redirected to the post-test Qualtrics survey where they entered demographic and education history. Participants took about 45-60 minutes to complete the entire experiment. See **Figure 3** for a schematic of the experiment.

Figure 3.

Part 1



Windmill rotates	Encoding Stimulus	Mask	Target Stimulus
120 frames	3.5 seconds	0.5 seconds	6.0 seconds

Results

Task Performance

In *Table 1* below, accuracy is presented as a function of expertise, rotation angle, condition, and chunk manipulation. Consistent with Stieff et al. (2020) both novice and naive participants had the greatest accuracy in detecting identical target stimuli. Also, both groups had the most difficulty in detecting chunk-maintained ball swaps. Overall, both groups had better accuracy with color change trials than swap trials, with the highest accuracy in trials where the color change disrupted a chunk.

Across all match and mismatch trials, novice participants performed better than naïve participants. Likewise, novices performed better than naive participants on average in Paper Folding and Cube Comparison Tasks.

Table 1. Mean accuracy for change-detection, Paper Folding, and Cube Comparison tasks with standard error in parentheses

	Novices *Mean(SE)*		Naives *Mean(SE)*	
Condition	10 degree	120 degree	10 degree	120 degree
Identical	0.88(0.01)	0.74(0.01)	0.85(0.01)	0.71(0.01)
Chunk-Changed Color Change	0.77 (0.02)	0.79(0.02)	0.71(0.03)	0.73(0.02)
Chunk-Changed Swap	0.73(0.02)	0.70(0.03)	0.59(0.03)	0.61(0.03)
Chunk-Maintained Color Change	0.76(0.02)	0.72(0.03)	0.59(0.03)	0.65(0.03)
Chunk-Maintained Swap	0.54(0.03)	0.40(0.03)	0.47(0.03)	0.34(0.03)
Paper Folding Test	11.80 (0.77)		8.29(0.85)	
Cube Comparisons	13.61 (2.31)		9.27 (2.32)	

Correlations with Spatial Ability Measures

We predicted that spatial ability would be more predictive of color swaps than color changes because a swap is a spatial transformation, whereas a color change is a visual transformation. In order to assess the relation between spatial ability and change detection sensitivity, we performed correlations of change detection sensitivity for color swaps and color changes with Paper Folding (M=10.05, SD=5.46) and Cube Comparison (M=11.44, SD=14.84) tests. The correlation table is shown in *Table 2*.

Table 2

Correlation table for the tasks and varying trial mismatch conditions in Part 1

****Note: r(p value)***

	Color Change Detection	Swap Detection	Paper Folding	Cube Comparisons
Color Change Detection	-	-	-	-
Swap Detection	0.21(0.06)	-	-	-
Paper Folding	0.13(0.24)	0.39(<0.01)	-	-
Cube Comparisons	0.05(0.65)	0.27(0.01)	0.58(<0.01)	-
Spatial Ability	0.10(0.36)	0.37(<0.01)	0.89(<0.01)	0.89(<0.01)

The two spatial ability measures, Paper Folding and Cube Comparisons, were highly correlated with one another, $r(80) = 0.58$, $p < 0.01$. Therefore, we combined the two measures into a single spatial ability score by computing the average of the z-transformed Paper Folding (PF) and Cube Comparison (CC) scores (z-transformation: scores for Paper Folding and Cube Comparison were averaged and centered to zero and then scaled to have a standard deviation of one). Color change detection was not significantly correlated with Paper Folding $r(80) = 0.13$, $p = 0.24$, Cube Comparisons $r(80) = 0.05$, $p = 0.65$, or Spatial Ability $r(80) = 0.10$, $p = 0.36$.

However, we see that swap detection has a small correlation with Cube Comparisons $r(80) = 0.27$, $p = 0.01$. Likewise, swap detection has a medium-sized correlation with both Paper Folding $r(80) = 0.39$, $p < 0.01$ and Spatial Ability $r(80) = 0.37$, $p < 0.01$ (cf. Cohen, 1988). Interestingly, color change detection was not significantly correlated with swap detection $r(80) = 0.21$ $p = 0.06$.

Overall, our analysis indicates that color swaps and color changes are distinct transformations. As predicted, swaps depend more on spatial ability than color changes. Additionally, while color changes aren't spatial in nature, they prove to be easier detections due to its being a visual transformation.

Linear Mixed Model

In order to test the effects of expertise, rotation angle, presence of change, condition, chunk manipulation, spatial ability, and their interactions on sensitivity to a change, a linear mixed model was created and fitted to the data. The model was implemented using R (R Core Team, 2021) and lme4 (Bates, et. al., 2015). The fixed factors in the model were expertise, rotation angle, presence of change, type of change (color change or swap), chunk manipulation, and spatial ability. Subject was treated as a random factor. *Table 3* indicates a summary of the estimates and p-values of the best-fit model.

Table 3. Coefficients table for linear mixed model

Fixed Effects	Estimates	Standard Error	χ^2	p-value	95% CI
Expertise	0.06	0.02	11.81	0.0006 ***	[0.03, 0.10]
Rotation	-0.05	0.01	15.50	<0.0001 ***	[-0.07, -0.03]
Presence of Change	-0.16	0.02	71.25	<0.0001 ***	[-0.20, -0.12]

Swap*	-0.16	0.01	133.21	<0.0001 ***	[-.19, -0.14]
Chunk Manipulation	0.15	0.01	106.24	<0.0001 ***	[0.12, 0.17]
Spatial Ability	0.02	0.01	7.52	0.01 **	[-0.01, 0.04]
Swap x Spatial Ability	0.04	0.01	5.86	0.02 *	[0.01, 0.06]

* Note: “Swap” indicates the difference in performance between a color swap and a color change.

As predicted, novices were more able to detect changes in molecular models $\chi^2(1) = 11.81$, $p < 0.01$ than naive participants. Across both novice and naive participants, a minimal (10 degree) angle change detection resulted in better accuracy than a 120 degree rotation $\chi^2(1) = 15.50$ $p < 0.01$, as predicted. This indicates that novices perform with higher accuracy on change detection with and without rotation, independent of spatial ability, suggesting an additional benefit to having studied organic chemistry. Participants also had better accuracy in trials that were the same as the base stimuli $\chi^2(1) = 71.25$, $p < 0.01$ than for change trials. Within change trials participants were more accurate in detecting color changes than color swaps $\chi^2(1) = 133.21$, $p < 0.01$, and in detecting changes that disrupted chunks rather than maintained them $\chi^2(1) = 106.24$, $p < 0.01$.

Furthermore, the model indicates that higher spatial ability (measured as a composite score on the Paper Folding and Cube Comparison tests) increases accuracy $\chi^2(1) = 7.52$, $p < 0.01$. Finally, we examined whether there was an interaction between type of change trials (swap vs. color change) and spatial ability and found that the difference between swaps and color changes was smaller for participants of higher spatial ability than for participants with lower spatial ability. This interaction also reflects the different correlations between spatial ability and swaps vs color changes, reported above.

Discussion

Our results show clearly that novice students did better than naive students in both the change detection tasks and on the spatial ability measures (Paper Folding and Cube Comparisons). This suggests that organic chemistry education does have an increased effect on performance. However, this claim does not help establish causality; rather, spatial ability and expertise are correlated. We have insufficient evidence to determine that an individual's chemistry education actively causes them to have increased accuracy in change detection and spatial ability tasks. Rather, it could be that only those with higher spatial ability choose to pursue STEM fields to begin with.

Furthermore, when analyzing novices and naive participants' performances, we notice that novices do better than naives even independently of spatial ability. This indicates an effect of domain knowledge over and above the differences in spatial ability between the two groups. This trend may even suggest that novices are able to recognize familiar domain-structures, like the ball-and-stick molecular models used in the change detection paradigm, and manipulate them as an advantage over those who do not study chemistry in their undergraduate education. Novice participants' increased accuracy could even suggest that students are starting to accumulate and practice expert-level strategies. However, we must acknowledge that beyond the familiar ball-and-stick models, the stimuli used did not in fact have chemically-relevant "chunks", or ligands commonly seen as part of larger molecules. Future improvements to this research could include surveying the specific strategies of participants' to better understand the level of conscious familiarity and recognition of chemistry-specific structures. These uncertainties limit us in hypothesizing specific strategies participants' may have utilized.

Lastly, our results indicate that within the change detection paradigm, color swaps in fact, more spatial transformations than color changes. Swaps, which involve the switching of two objects within a structure, involve movement within a plane (a spatial transformation), requiring individuals to recruit an ability to manage multiple changing pieces of information. However, color changes require individuals to instead recognize a change of color patterns for a single object within the larger structure. Our evidence supports this theoretical distinction as color changes and color swaps are not highly correlated with one another. These results also suggest that there may exist a distinction between visual and spatial working memory. The finding that color swaps are better at detecting spatial ability than color changes can have great implications in spatial ability research. Thus, future research should incorporate these swap transformations as a testable and detectable spatial ability measure in their paradigms.

Overall, the understanding of how spatial ability can be tested with the domain of chemistry has great pedagogical implications. With a better understanding of the types of changes students struggle with and which ones come at ease to STEM learners, we can better cater our online learning programs and in-person teaching strategies to better practice these specific transformations. While we do not fully understand the direction of causality of whether higher spatial ability leads to success in these change detection paradigms or whether it is solely the effect of chemistry education, we now have further evidence supporting that higher spatial ability does indeed lead to success in STEM. Our research paves the way in new research that can now test the bounds of whether this spatial ability is fluid and whether it can be strengthened with practice through effective educational methods.

Acknowledgements

We thank UCSB Hegarty Spatial Lab research assistants: Scott Matsubara, Joanne Kim, and Tyler Chia for their consultation and assistance in data analysis, as well as Lab Manager, Daniel Buonauro for administrative support. Additionally, I'd like to thank Chuanxiuyue He and Dr. Mary Hegarty for their guidance and support every step of the way.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*(4), 487–502. https://doi.org/10.1037/a0016797

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Ekstrom RB, French JW, Harman HH, Derman D. Manual for kit of factor-referenced cognitive tests. Princeton: Educational Testing Service; 1976.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: enhancing students' pattern recognition, structure extraction, and fluency. *Topics in cognitive science*, *2*(2), 285–305. https://doi.org/10.1111/j.1756-8765.2009.01053.x

Luck, S., & Vogel, E. (1997). The capacity of visual working memory for features and conjuctions. *Nature*, *390*(6657), 279-281. doi:10.1038/36846

Miller G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2). doi: 10.1037/h0043158.

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic bulletin & review*, *15*(1), 215–222. https://doi.org/10.3758/pbr.15.1.215

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*(3), 604–614. https://doi.org/10.1037/0022-0663.93.3.604

Stieff, M., Uttal, D. How Much Can Spatial Training Improve STEM Achievement?. *Educ Psychol Rev* 27, 607–615 (2015). https://doi.org/10.1007/s10648-015-9304-8

Stieff, M., Werner, S., DeSutter, D., Franconeri, S., & Hegarty, M. (2020). Visual chunking as a strategy for spatial thinking in STEM. *Cognitive research: principles and implications*, *5*(1), 18. https://doi.org/10.1186/s41235-020-00217-6

Super, D. E. & Bachrach, P.B. (1957). *Scientific Careers and Vocational Development Theory*. New York: Teachers College, Columbia University

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), 352–402. https://doi.org/10.1037/a0028446

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

Xu, Y., & Franconeri, S. (2015). Capacity for visual features in mental rotation. *Psychological Science*, *26*(8), 1241-1251. doi:10.1177/0956797615585002

Appendix A

Qualtrics Demographic Post-Experiment Questionnaire

This Appendix contains the questions presented to participants after the change detection task, paper folding, and cube comparisons task.

Questions:

1) Sex
2) Age
3) Major
4) Have you taken, or are currently taking, any college-level chemistry courses? If Yes, see Question 5. If No, proceed to Question 6.
5) In the box below, please list any college-level chemistry courses you are currently taking.
6) Are you looking to receive class credit for completing the study? If Yes, proceed to Question 7. If No, proceed to Question 8.
7) Please provide your name so that we can award you SONA class credit.
8) Please enter an email that you would like us to send a $10 Amazon Gift Code to.