University of California
Santa Barbara

Ability moderates alternate strategy use during the mental rotation task with molecule-like stimuli

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Psychological and Brain Sciences

by

Trevor J. Barrett

Committee in charge:

Professor Mary Hegarty, Chair
Professor Richard Mayer
Professor Daniel Montello
Professor Russel Revlin

September 2019

The Dissertation of Trevor J. Barrett is approved.

Professor Richard Mayer

Professor Daniel Montello

Professor Russel Revlin

Professor Mary Hegarty, Committee Chair

June 2019

Ability moderates alternate strategy use during the mental rotation task with molecule-like stimuli

Copyright © 2019

by

Trevor J. Barrett

To my loving family and friends. And to my steadfast champion and inspiration, Micaela Rae.

Acknowledgements

I would like to thank my advisor, committee members, and the department staff who offered critical advisement and patience throughout.

Curriculum Vitæ

Trevor J. Barrett

Education

2019	Ph.D. in Psychology, University of California, Santa Barbara.
2015	M.A. in Psychology, University of California, Santa Barbara.
2011	B.A. in Psychology University of California, Santa Barbara.

Publications

Barrett, T. J., & Hegarty, M. (2016). Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. *Computers in Human Behavior, 65*, 220-231.

McKenzie, G., Hegarty, M., Barrett, T. J., & Goodchild, M. (2016). Assessing the effectiveness of different visualizations for judgments of positional uncertainty. International *Journal of Geographical Information Science, 30(2)*, 221-239.

Barrett, T. J., Stull, A. T., Hsu, T. M., & Hegarty, M. (2015). Constrained interactivity for relating multiple representations in science: When virtual is better than real. *Computers and Education, 81*, 69-81.

Barrett, T. J., & Hegarty, M. (2014). Interaction design and the role of spatial ability in moderating virtual molecule manipulation performance. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 672-677). Cognitive Science Society Austin, TX.

McKenzie, G., Barrett, T. J., Hegarty, M., Goodchild, M., & Thompson, W. (2013). Assessing the effectiveness of visualizations for accurate judgments of geospatial uncertainty. Visually-Supported Reasoning with Uncertainty workshop at the Conference on Spatial Information Theory (COSIT '13). Scarborough, UK.

Barrett, T. J., & Hegarty, M. (2013). Interface design and spatial cognition: A case of virtual molecule manipulation. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 157-162). Austin, TX: Cognitive Science Society.

Stull, A. T., Barrett, T. J., & Hegarty, M. (2013). Usability of concrete and virtual models in chemistry instruction. *Computers in Human Behavior, 29(6)*, 2546-2556.

Stull, A.T., Barrett, T. J., & Hegarty, M. (2012). Design of a virtual reality system for the study of diagram use in organic chemistry. In *Proceedings of the Workshop on Technology-Enhanced Diagrams Research* (pp. 33-41).

Abstract

Ability moderates alternate strategy use during the mental rotation task with molecule-like stimuli

by

Trevor J. Barrett

Using the classic mental rotation task paradigm, I compared participants' performance with Shepard and Metzler (1971) cube figure stimuli and tetrahedron molecule-like stimuli. In Eperiment 1, I used the classic mental rotation task paradigm to compare performance with the original cube figure stimuli and molecule-like tetrahedron stimuli as well as axis of rotation – picture- vs. depth-plane. In Experiment 2, I tested the effect of display dimensionality – stereoscopic 3D versus standard mono, and also included additional trials to investigate practice effects. In Experiment 3, I looked at the moderating effect of ability on performance and strategy use. In all experiments, response time and error rate as well as slopes and intercepts were calculated to examine overall performance and individual differences. Specifically for trials with rotations about the horizontal depth-plane axis, the left and right elements did not move laterally across angular disparities. This allowed for a non-rotation strategy that involved comparing the relative locations of the adjacent elements, rather than utilizing a mental rotation strategy on rotations about the horizontal depth-plane axis. Individuals were assigned to strategy use classes based on both self-report and statistical clustering of individual slope coefficients (i.e. rotation rates). The moderating effect of ability on performance and strategy use was assessed three measures of ability: Cube Comparison, Paper Folding, and Abstract Reasoning. Extending Shepard's classic mental rotation task paradigm, I utilized desktop virtual reality software for stimuli presentation and data collection. On depth-plane rotations, the angular disparity effect (ADE) was significantly greater on cube figures compared to tetrahedrons, while similar slopes were observed on both stimuli types for picture plane rotation trials. I found general agreement between both assessments of strategy use— self-report and clustering on individual slope coefficients. Two divergent strategy classes emerged after repeated practice with molecule-like tetrahedron stimuli: a) a mental rotation strategy evidenced by a positive angular disparity effect, and b) a non-rotation strategy with slopes too small to consider mental rotation. Participants picked up on the non-rotation strategy about 70% of the time and discriminant analysis showed the classifier had a predictive error of 30% against self-reported strategy as labeled data. In general, rotators scored higher on Cube Comparison and the two strategy groups did not differ on Paper Folding or Abstract Reasoning. In terms of performance, higher Cube Comparison scores were predictive of larger slopes for tetrahedron stimuli, indicating use of a mental rotation approach for both stimuli types. High abstract reasoning scores predicted larger slopes for cube figure stimuli and flatter slopes for tetrahedron stimuli - indicating successful strategy switching. This indicated that higher performing participants were able to flexibly switch between an orientation-independent strategy for tetrahedron stimuli and a mental rotation for the cube figure stimuli. These findings improve our understanding of the interplay between ability and strategy for reasoning about spatial tasks with domain-specific stimuli. Importantly, I provide evidence that the molecule-like tetrahedron stimuli were not necessarily processed the same way as stimuli used by Shepard and Metzler (1971). Cube Comparison Test score predicted use of a mental rotation strategy and Abstract Reasoning predicted use of an alternate non-rotation strategy for the tetrahedron stimuli. Individual differences in ability are discussed in the context of spatial judgement tasks for organic chemistry instruction.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Spatial thinking skills are important in many scientific fields as well as in everyday life. For example, spatial thinking is critically involved in reasoning about geological formations or understanding molecular structures in a medicinal chemistry laboratory. Longitudinal research has shown spatial ability to be uniquely predictive of achievement in the sciences (Wai, Lubinski, & Benbow, 2009). A recent meta-analyses provided strong evidence that spatial skills are trainable and are at least somewhat transferable (D. H. Uttal et al., 2013). Mental rotation, or the visual inspection and mental simulation of an object's rotation in space (Hegarty & Waller, 2005), is a spatial skill that has been researched in the context of organic chemistry education (Stieff, 2011; Stieff & Raje, 2010; Stieff, Ryu, Dixon, & Hegarty, 2012). The sub-field of stereo-chemistry involves the study of the three-dimensional spatial arrangement of molecules and their electronegativity. Biologically relevant molecules are often chiral, meaning that their spatial configuration has "handedness" – asymmetry such that its mirror image is non-superimposable. These molecules are called *enantiomeric* pairs and, critically, often differ greatly in their biological activity. Chemistry students conducting organic synthesis learn to control the structural arrangement of chiral molecules by using enantiomer-specific reactions. Understanding concepts relating to chirality, such as being able to discriminate between chemical enantiomers is a crucial, however notoriously challenging foundational concept for many students in organic chemistry.

The task of enantiomer discrimination in organic chemistry— determining whether two molecules are non-superimposable mirror images, is analogous to the Shepard and Metzler (1971) experimental task where participants judged whether two stimuli were identical but rotated or were enantiomeric mirror-reflected figures. However, these surface-level experimental task similarities may mask important differences between stimuli typically used in chemistry instruction and those used in spatial cognition research (Shepard & Metzler, 1971). Mental rotation tests typically utilize stimuli consisting of three-dimensional structures with ten concatenated cubes that form two "arms and legs" at the terminal ends of the object. Overall, the stimuli used to test mental rotation ability differ significantly from molecular models used in instruction, in terms of the overall three-dimensional spatial structure, visual access and occlusion, as well as color coding of domain-specific information (i.e. atoms). In the following chapters, I will argue that these crucial differences– along with constraints set up by test characteristics pose potentially critical implications regarding spatial skills training for STEM education.

In Chapter II, I begin by reviewing classic chronometric studies of mental rotation. Here I clarify and define the mental rotation process versus mental rotation as an experimental task paradigm or psychometric ability measure. I summarize previous research on mental rotation by looking at the different types of tasks employed. Further, I examine previous mental rotation studies that tested variations in stimuli type, complexity, dimensionality, and axis of rotation. In addition to examinations of task and stimuli manipulations, I also review the importance of Mental Rotation Tests as a measure of spatial ability. Previous research has shown there to be alternate strategies available on tests such as the Vandenberg and Kuse MRT because they were designed differently than the traditional Shepard and Metzler mental rotation paradigm. I review studies that have previously examined alternate cognitive strategies during mental rotation and other spatial tasks. Finally, I review research at the intersection of spatial thinking and chemistry instruction.

In Chapter III, I used the classic mental rotation task paradigm to compare performance with the original cube figure stimuli and molecule-like tetrahedron stimuli. Here I extended the Shepard and Metzler (1971) experimental task paradigm with realistic lighting and shadows rendered on a stereoscopic display. Participants completed the mental rotation task with molecule-like tetrahedron figures as well as traditional Shepard and Metzler (1971) cube figures for both picture plane rotations and depth-plane rotations about the horizontal axis. The findings showed that on aggregate, reasoning about both stimuli types appeared to invoke mental rotation – evidenced by positive angular disparity effects for both stimuli types. Angular disparity effect (ADE) is defined as the positive linear effect of angle observed on response time, and for all intents and purposes, has been offered as a behavioral signature of the mental rotation process. Notably, a significantly smaller slope was observed for the tetrahedron stimuli compared to the cube figures, specifically on depth axis rotation trials. Closer analysis of individual differences indicated that a subset of individuals was likely utilizing an alternate non-rotation strategy only available on depth-axis rotation trials. Overall, Experiment 1 provided a controlled comparison between reasoning about spatial stimuli common in experimental psychology research and tetrahedral ball-and-stick models frequently reasoned about in organic chemistry instruction. Next, I attempted to distinguish tetrahedron rotators versus non-rotators based on their slopes.

In chapter IV, I report an experiment addressed multiple research questions. One question was whether providing 3D stereo viewing affected performance. To test this, I directly compared stereo and mono display formats between subjects in order to rule out display dimensionality as a confounding variable. A primary focus of Experiment 2 was to attempt to distinguish cognitive strategy use across stimuli type and level of practice. Trials were presented in counterbalanced blocks of each stimuli type to minimize ordering effects. Self-reports of strategy as well as statistical clustering on participants' slope coefficients indicated that a subset of participants utilized a mental rotation strategy for both stimuli types while a larger group of participants utilized an alternate non-rotation cognitive strategy for the tetrahedral stimuli. This finding indicated that I were able to distinguish between participants' cognitive strategy use based on behavioral data. Next, I wanted to see how different measures of ability moderated alternate strategy use and performance.

In chapter V, I was interested in whether ability measures were predictive of performance and/or cognitive strategy use. Three measures of ability were administered: Cube Comparison, Paper Folding, and Abstract Reasoning (Ekstrom, Dermen, & Harman, 1976). For this experiment, I report the findings of a full Bayesian analysis to predict strategy use and performance depending on the three measures of ability. Participants with higher Cube Comparison scores were more likely to be classified as rotators. Rotators were a class of participants that stuck with the mental rotation approach to solving both the standard cube figure stimuli trials as well as the molecule-like tetrahedron trials. Abstract Reasoning was predictive of alternate strategy use along with faster and more accurate performance compared to those who scored lower on the Abstract Reasoning test. Further, I found that people with higher Cube Comparison scores were more likely to be assigned to the rotator strategy classification. These findings suggest that cognitive strategy use is influenced by various abilities that are not necessarily spatial in nature.

Finally in Chapter VI, I draw theoretical conclusions and discuss the implications and applications of the present research. This work highlights how seemingly minor alterations to experimental task or stimulus can have major impact on task performance due to individual differences in strategy use. I find that individuals with a high propensity for mental rotation, an effective strategy for the Cube Comparison test, tended to utilize that strategy despite there being an alternative strategy. Did rotators know about the alternate strategy, yet chose to continue using mental rotation? Or, were participants too fixated on mental rotation in the context of the experiment to notice the existence of the alternate strategy? Further, I suggest more broadly that nominally non-spatial abilities such as abstract reasoning and pattern recognition likely played an important role in what we think of as well-defined spatial abilities (mental rotation). If purported pen-and-paper psychometric measures of mental rotation afford multiple strategies, then the measure in question will necessarily be lacking in ecological validity. Limitations inherent to the two-choice forced response time task are discussed. I end on how our theoretical conclusions offer future directions for collaboration between spatial thinking researchers, STEM educators, and policymakers.

Chapter 2

Background

Chronometric studies of mental rotation

Most psychologists know mental rotation as the task of imagining how a two- or three-dimensional object would look if rotated away from its original upright position. In his keynote address "How a cognitive psychologist came to seek universal laws", Shepard (2004) described a distant memory at Stanford wherein he "experienced a spontaneous hypnopompic image of three-dimensional objects majestically turning in space." This daydream inspired the stimuli and experimental design that made it onto the cover of *Science* three years later (Shepard & Metzler, 1971). In Shepard's words,

> Mental rotation represents a more voluntary [versus apparent motion] and cognitively effortful case. In it, the time to determine that two displayed objects are of the same shape similarly increases linearly with the extent of the kinematically simplest rigid motion required to bring one object into the orientation of another in the three-dimensional space. Here, too, this simplest motion is affected by any symmetries of the object. Despite the

> absence of external support, the object is sequentially represented in positions corresponding to successive points along this particular simplest connecting path in the abstract space of possible positions. This is demonstrated by how the speed of the (accurate) discriminative response to a visual probe presented during the mental operation systematically varies as a function of the time and relative orientation of the presented probe. (Shepard, 2004, pp. 7–8)

The mental rotation experiments are foundational in experimental psychology research as they provide a paradigm for investigating cognition by measuring mental imagery. There were two initial types of mental rotation tasks that Shepard's group used. The first type simultaneously displayed perspective views of two three-dimensional objects, as in Figure 2.1, and measured the time for participants to determine whether the two objects, though differing in their orientations, were either the identical or it's mirror image (i.e. enantiomer). Other mental rotation experiments such as those by Cooper and colleagues used singular alphanumeric numbers and letters displayed at differing orientations and were recalled from memory (Cooper, 1975; Cooper & Podgorny, 1976). In general, the estimated rates of mental rotation were always lower for the simultaneous presentation tasks (20-100 deg/s) than for tasks comparing a 2D stimulus with a stored internal representation recalled from memory (300-600 deg/s). Critically, response time increased linearly as a function of angular disparity under both experimental task paradigms. This early work on mental rotation was important because it demonstrated that transformations of mental imagery could be functional in thinking. Another advantage of the mental rotation paradigm was that they did not depend on verbal self-reports from participants nor explicit instructions to use imagistic strategies during the task. Regarding the three experiments presented in the following chapters, I used the original experimental design:

Figure 2.1: Example of original cube figure stimuli from Shepard and Metzler (1971). The pictured 'identical' trial involves a rotation about the picture-plane axis to align the two objects.

a two-choice forced response task where participants made same/different judgments for each pair of simultaneously displayed stimuli at various angular disparities.

Importance of mental rotation as measure of spatial ability

In addition to being utilized as an experimental task, mental rotation has been critical in demonstrating the functional importance of transformations of mental imagery in spatial thinking. Mental rotation has been important in studies of psychometric intelligence as one of the more widely used measures of spatial ability (Carroll, 1993). Mental rotation is also known to produce some of the largest and widely replicated sex differences in the cognitive psychology literature has been suggested to be decreasing magnitude over time (Daniel Voyer et al., 1995). The Vandenberg and Kuse Mental Rotation Test (Perceptual & Motor Skills, 4, 599-604, 1978) is one of the most frequently cited psychometric measures of spatial ability (Hegarty, 2018). Notably, there is a long history of reported sex differences on the (**???**) Mental Rotation Test in favor of men (Linn & Petersen, 1985), with meta-analyses indicating effect sizes between 0.7 and 1.0 standard deviation that appears to be narrowing over time (D Voyer et al., 1995).

Another measure of spatial ability commonly used is the Purdue spatial visualization test: Visualization of Rotations. A 2013 meta-analysis (Maeda & Yoon, 2013) explored whether the widely observed sex differences in mental rotation ability on this test resulted from procedural artifacts from methodological choices. Overall, they found that the male advantage on spatial ability measured by the PSVT:R was related. Importantly, the researchers found the sex difference increased when more stringent time limits were administered.

It is important to note that not all measures of spatial ability are necessarily solved using pure mental rotation. As an early example, Just & Carpenter (1985) demonstrated that The Cube Comparisons Test was measure of spatial ability that could be solved using either a mental rotation approach or alternate non-rotation strategies.

Cognitive strategies in spatial measures

Spatial cognition researchers have offered that differences in performance on spatial ability measures may be explained at least in part by the existence and utilization of an alternate cognitive strategies that obviate mental rotation for certain items on the Vandenberg and Kuse Mental Rotation Test (Geiser, Lehmann, & Eid, 2006; Gluck & Fitting, 2003; Hegarty, 2018). Geiser et al. (2006) describe this strategy as invoking an alternate non-rotation process rather than mental rotation to efficiently eliminate false answer choices. An analysis of the MRT items reveals that many of the response foil choices can be eliminated by simply noticing that the end-arms are perpendicular rather than parallel. Some foils have a different number of end-arm cubes and can also be eliminated without using mental rotation. Understanding that mental rotation is not necessary in all cases allows the test taker to complete more of the multiple-choice items.

Geiser et al. (2006) conducted a multigroup analysis on the MRT and revealed some participants were using a strategy that could be classified as "rotators", and other participants were using a "non-rotator" strategy. This finding was important as it indicated unintuitively that several items on the MRT don't require a mental rotation approach to most efficiently arrive at the solution. Geiser et al. (2006) also found that there was a group of people that had extremely poor performance, possibly arising from strategic indecision.

Gardony et al. (2017) predicted that the Mental Rotation Test invokes both motor simulation and other analytic cognitive strategies that depend on the visualization representation stored in visual working memory. They suggest that performance on the Mental Rotation Test is not purely a measure of mental rotation as a cognitive process but instead also involves flexibly exploiting a working memory intensive analytic strategy depending on task difficulty.

Boone and Hegarty (2017) offered that the sex difference may be driven by male's propensity to exploit alternate non-rotation strategy when available on the subset of items with different numbers of end-arm cubes or whether the end-arms were parallel or perpendicular. They suggest that non-rotation strategies such as counting the number of end-arm cubes or exploiting asymmetries may drive individual differences in performance. Previous work by Hegarty (2018) found that most participants reported using both analytic and mental rotation strategies. The most successful students tended to use an analytic strategy to eliminate answer choices, thereby reducing mental imagery time. Notably, women were more likely to report using non-rotation strategies that were not necessarily advantageous, such as counting the number of cubes.

These findings suggest that when speeded mental rotation tasks are used for personnel selection batteries, a large group of people will stick with using a mental rotation approach despite there being more efficient strategies available on certain items. The MRT consists of two 10 item blocks and participants are given 3 minutes to complete each block. This is in contrast with how Shepard (1974) described his participants: highly trained on the mental rotation task and they often completed thousands of trials across multiple experiments. These results offer converging evidence that cognitive strategy use may be a domain-general principle that underlies spatial ability or other aptitudes more broadly. In the context of this the present research, I assigned participants to strategy class based on their behavior and self-reported strategy use.

Spatial thinking and STEM education

For more than a decade, education researchers (Shea, Lubinski, & Benbow, 2001; Webb, Lubinski, & Benbow, 2007) and policymakers (National Research Council, 2006) have highlighted the importance of spatial thinking in science, technology, engineering, and mathematics (STEM) fields. Much enthusiasm has stemmed from evidence that spatial abilities are predictive of higher achievement in STEM disciplines (Shea et al., 2001; Wai et al., 2009). Cognitive psychologists have studied spatial thinking in STEM domains like chemistry (Wu & Shah, 2004) and engineering (Sorby, Casey, Veurink, & Dulaney, 2013), for example. The prospect of spatial cognition research impacting STEM education has been bolstered by strong evidence demonstrating the malleability of spatial thinking skills through training interventions. The Uttal et al. (2013) meta-analysis of spatial skills showed that substantial gains in spatial reasoning performance can be had through training. Further, their analysis also offered evidence that spatial skills persist after training and are at least somewhat transferable across a variety of stimuli. The authors suggest that spatial skills training could potentially lead to a significant increase in STEM participation and achievement.

In STEM domains such as chemistry, spatial problem solving often involves the task of resolving molecular identity from color- or component-bindings, rather than by structure alone as in the original mental rotation stimuli. Discriminating between enantiomers is an example of reasoning about the concept of chirality in chemistry. In general, most representations of organic molecules share similar geometries that only differ in the configuration of their component parts. Molecular models with tetrahedral geometry used in organic chemistry instruction have a set spatial structure and are only identifiable by their color-bindings. Tetrahedral molecules can also be chiral– where the mirror image of the molecule is a non-superimposable enantiomer. In contrast to the stimuli used by Shepard and Metzler (1971), the identity of tetrahedral molecules are defined by the arrangement of bound components in a fixed tetrahedral configuration rather than by the spatial structure itself. In the original mental rotation task paradigm, the identity of the stimuli was based purely on how each block of cube-figures were concatenated in space, with its enantiomeric pair necessarily being its mirror image. In contrast, reasoning about chirality in introductory organic chemistry involves comparing spatial structures with identical geometries but differ in the configuration of their component parts.

Chemists deal with spatial complexity by employing a rich spatial language that encodes a finite set of molecular geometries and components. One analytic strategy for enantiomer discrimination commonly taught in introductory organic chemistry involves identifying then counting components ordered by their chemical reactivity. A molecule with clockwise ordering is necessarily the enantiomeric pair of a molecule with counter-clockwise ordering of the same components. Another non-rotation strategy involves identifying a plane of symmetry and comparing components across the plane.

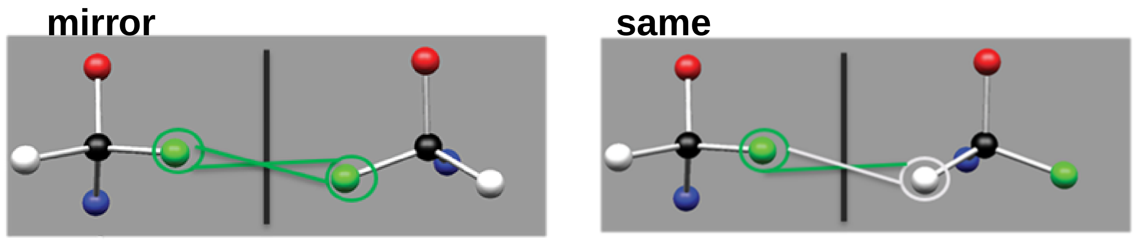

Figure 2.2: Example of same and different trials in a mental rotation task with molecule-like stimuli

Because most molecular visualizations are bound by domain conventions, regularities in geometric structure are predictable. In the context of organic chemistry instruction, structural regularity of molecules combined with standard conventions for visualization enable use of analytic strategies that are unavailable with objects studied in typical mental rotation experiments.

Domain specific coding and complexity reduction schemes are utilized by expert chemists, so the role of spatial abilities such as mental rotation is important when first learning about spatial concepts. When students' knowledge of nomenclature and analytic strategies have not yet developed, using mental imagery can be an effective method for reasoning about molecular structures. Color bindings and structural regularities of organic molecules afford use of analytic strategies that are unavailable for the original cube figure stimuli. This subtle but important qualitative difference between the stimuli studied in classic research on mental rotation versus objects frequently reasoned about in a STEM domain is a primary focus of this research. Understanding the relationship between mental rotation ability and performing an analogous task in the context of STEM domains is critical if we are to generalize findings and offer implications for improving STEM learning.

The idea that spatial skills training could improve achievement in STEM has been explored with optimism (Uttal & Cohen, 2012; Uttal et al., 2013). However there has been little empirical evidence establishing a causal link between spatial skills training and achievement. A notable example of spatial skills training has been in the case of engineering education (Sorby, 2009; Sorby et al., 2013). Sorby et al. trained various general spatial transformations and found gains in achievement outcomes of engineering students. Earlier work in the domain of chemistry by Small and Morton (1983) reported that an intervention using physical models led to small gains in course achievement. More recently, Miller and Halpern (2013) workshop utilizing materials developed by Sorby and found that improvement on spatial ability measures were observed but found weak support for an impact on achievement.

Stieff and Uttal (2015) present a realistic view of the evidence so far and offer cautious optimism regarding what can be expected from spatial interventions for STEM achievement. In addition to pointing out that only weak evidence has been presented thus far, they highlight the methodological issues associated with self-selection and quasi-experimental methods that are inherently difficult to avoid in classroom interventions. More critical is the fact that while showing intervention efficacy is important, little work has been done to offer explanations or mechanisms of spatial skills gains regarding whether training causally impacts achievement. Well controlled empirical studies are needed for developing stronger theories of expected efficacy of training interventions and for understanding the degree to which training is transferable to relevant domains.

Stieff and Raje (2010) showed that organic chemistry experts tend not to use mental rotation, and argued instead that experts are able to exploit more generalizable patterns of symmetry inherent to chemical structures. Studying a group of organic chemistry students, Stieff (2011) showed that students used imagistic reasoning strategies for completing translations between various molecular representations in simple contexts but had difficulty integrating self-generated diagrammatic information for more complex problems. Further, Stieff et al. (2012) showed that spatial ability was beneficial in certain contexts for organic chemistry students that utilized the mental rotation approach, however experts were not benefited by having higher spatial ability because they exploited alternate non-rotation strategies. Hegarty, Stieff and Dixon (2013) conducted an intervention study for organic chemistry instruction and administrated training that emphasized imagistic strategies (e.g. mental rotation), analytic non-rotation problem solving strategies, or a combination of both imagistic and analytic strategies. They found that combined instruction with imagistic and analytic strategies eliminated the sex difference in course achievement. After the combined intervention, imagistic strategies were less commonly employed, indicating a switch in solution strategy from imagistic to analytic over the course of instruction. The authors suggest that imagistic mental models are associated with novelty, and that students adopt more domain-specific heuristics or rules-based approaches when made apparent.

In the following chapters I present three experiments to address whether a common cognitive process is evoked during the mental rotation task with both molecule-like tetrahedrons and cube figure stimuli.

Based on the previous literature, I predicted that mental rotation tasks of molecule-like tetrahedron stimuli may be accomplished by strategies other than mental rotation, depending on task characteristics such as axis of rotation. In Experiment 1, I manipulated stimulus type - cube figure vs. tetrahedrons and axis of rotation – picture vs. depth-plane. Here I demonstrated that some participants likely used an alternate strategy for tetrahedron trials rotated about the depth-plane axis. In Experiment 2, I classified participants based on their strategy questionnaire responses as well as statistical clustering on individual response time slopes. After selecting the strategy classification model with the best fit, I determined how strategy use affected performance. For each strategy class, I compared response times and error rates and with special attention to angular disparity effects. Strategy use was expected to affect performance, so I compared the behavior of each strategy classification group to highlight the strengths and weaknesses of each strategy. Next, I examined the relationship between strategy use and spatial ability. Experiment 3 replicated the previous experiment with the addition of a spatial test battery consisting of Cube Comparison, Paper Folding, and Abstract Reasoning Tests. To explore potential differential impacts of various ability measures across stimuli types I conducted full Bayesian regression analyses that indicated Cube Comparison was predictive of the mental rotation strategy, while Abstract Reasoning was associated with having flatter response time slopes and better accuracy for tetrahedron trials.

Chapter 3

Experiment 1: Stimulus and task characteristics

Introduction

Experiment 1 provided a controlled comparison between reasoning about stimuli common to psychological research and objects reasoned about in the domain of organic chemistry. I compared performance with traditional Shepard and Metzler (1971) mental rotation stimuli and stimuli relevant to a STEM domain- organic chemistry. The task was designed after Shepard and Metzler's (1971) simultaneous presentation paradigm. The original mental rotation task was extended by displaying high quality renderings of stimuli with a point light source and shading viewed through stereo glasses.

Based on previous research, I predicted that if reasoning about both stimuli types invokes the same process, the signature of mental rotation should be apparent for both. Specifically, if participants are using mental rotation, response time should increase linearly with angular disparity for both cube figure and tetrahedron stimuli. If response time does not increase linearly with angular disparity, it suggests that another process is being utilized besides mental rotation. Differences in strategy were expected to manifest as effects on slope– rather than on intercept, so analysis of intercepts was included as a matter of convention and for completeness. I predicted that participants may pick up on a non-rotation strategy such as comparing the colors of adjacent stimuli components. If a participant picks up an alternate strategy for the tetrahedron depth-plane rotation trials, then we would expect to see flat response time slopes compared to cube figure trials.

Method

Design.

Experiment 1 had a two (stimulus type: cube figure or tetrahedron) by two (axis of rotation: horizontal depth or picture plane) by six (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) by two (identity: identical or mirror) within-subjects design (see Figure 3.1). The dependent measures were response time (on accurate 'same' trials) and error rate.

Participants.

Thirty-six college students (20 Female) (age: $M = 18.8$, $SD = 1.8$) from the psychology subject pool at a research university participated in the study in return for course credit. None of the participants had previously studied organic chemistry. All participants had normal or corrected-to-normal vision. The order in which participants completed the experiment was randomly assigned. 18 participants completed the cube figure trials first, and 16 completed the tetrahedron trials first. All trials were administered within-subjects.

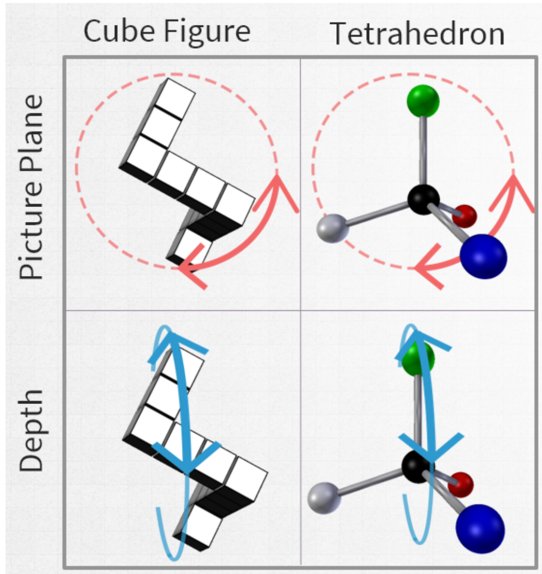

Figure 3.1: Stimuli and axes of rotation used in Experiment 1.

Materials.

The study materials included an informed consent form, instructions, and four practice trials with feedback. The 3D model stimuli were built using open source CAD software (Blender). A total of 384 trials were constructed (2 Stimulus x 2 Axis x 2 Identity [mirror or same] x 12 Angle) with four repetitions. Stimuli were displayed at various orientations within levels of angular disparity in order to mitigate item and environment-based alignment effects and object recognition. The four different stimulus-axis conditions are depicted in Figure 3.1. For the trials, stimuli were presented in pairs and differed in angular disparity from 30° to 330° in 30° increments. The number of 180° trials was doubled in order to keep observation counts balanced when opposite angles were collapsed during analysis. Half of the trials were rotated in the depth-plane about the horizontal axis and the other half involved picture-plane rotations. Half of the trials displayed identical stimuli, and half displayed enantiomeric mirror images. For the cube figure stimuli, four 3D models were built based on the original Shepard and Metzler figures and consisted of

10 concatenated white cube figures with black edges as shown. Four tetrahedral 3D models were built to resemble ball-and-stick chemistry models and had roughly equivalent dimensions as the cube figure stimuli. See Figure 3.4 for example trials.

Apparatus.

Stimuli were presented in stereo 3D using Nvidia 3D Vision Wireless Glasses (Santa Clara, CA) at 1920x1080 resolution. WorldViz Vizard (Santa Barbara, CA) virtual reality software was used for stimulus presentation and data collection. Trials were rendered using perspective and displayed shadows from a single light source.

Procedure.

Groups of up to three individuals at a time participated in a windowless laboratory room and were supplied with noise canceling headphones to minimize key press noises from other participants. After providing informed consent, participants were instructed on the use of the 3D glasses and verified with the experimenter that the image displayed was fusing properly for them and could be seen clearly. Next, the instructions were read aloud by the experimenter while participants followed along with written instructions displayed on the monitor:

> Your goal is to decide whether two objects are either identical (same) or mirror images of each other (different). Pairs of objects will be displayed at different orientations, and each pair will always be identical or mirror images. The objects in Figure 3.2 are the same given in four different positions. Notice how shadows may make altered appearances even though the objects are identical. Look at each to satisfy yourself that they are identical and only presented at

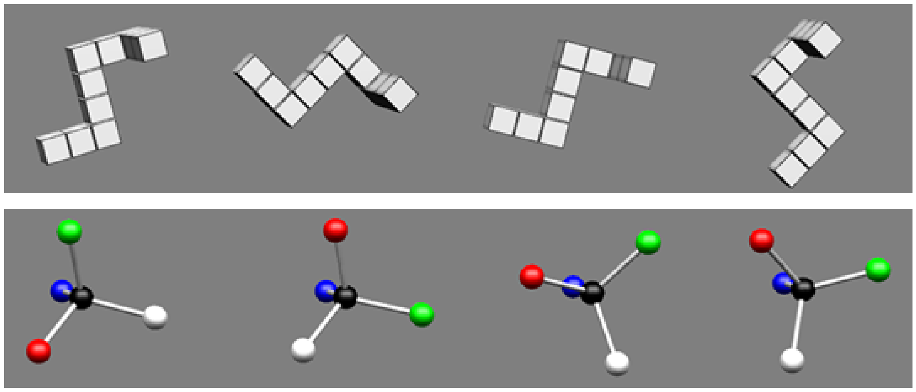

Figure 3.2: Examples of cube figure and tetrahedron stimuli displayed at various angular disparities.

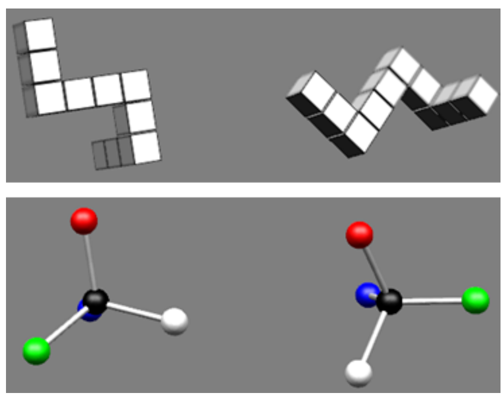

Figure 3.3: Sample of two 'different' trials that were shown to participants during instruction.

> different angles. Figure 3.3 shows two new objects. They cannot be made to match each other. Satisfy yourself that they are indeed different objects. Please confirm with the experimenter that you understand the task, or ask for clarification if needed. Respond by using the blue keys labeled S (same) & D (different). Please make your judgements as quickly *and* accurately as possible.

At the beginning of the study, participants completed four practice trials with feedback from and were instructed to work as quickly and accurately as possible through the self-paced experimental trials. After each block of trials, participants were given a break before moving on to the next block of trials with the other stimulus type or axis of rotation. Example trials are shown in Figure 3.4. Half of the participants completed the cube figure trials first and the other half completed the tetrahedron trials first. Separate instructions between each block were presented. All participants completed the picture-plane rotation block followed by the depth-plane rotation block.

Results and Discussion

Data screening

The following data screening procedures were applied in each of the following experiments. First, data from eight participants were eliminated for having lower than 60% overall accuracy rates. Screening for fast and careless respondents showed that there were no participants that had less than .5s response times on more than 20% of total trials. Responses faster than .5s (0.35% of total trials) were not included in the analysis. Outlying slow responses greater than 3.5 *SD* from an individual's mean for each cell in the 6 (angle) x 2 (stimulus) x 2 (axis) factorial design were not analyzed and accounted for 1.02% total trials. Data from the remaining 28 participants (15 female) was used for repeated measures ANOVAs with a 2 (stimulus type: cube figure, tetrahedron) x 2 (axis of rotation: depth, picture) x 6 (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) design. Response time and accuracy and their derived slopes were the dependent variables. Violations of the assumption of sphericity were addressed by reporting corrected Greenhouse-Geiser estimates.

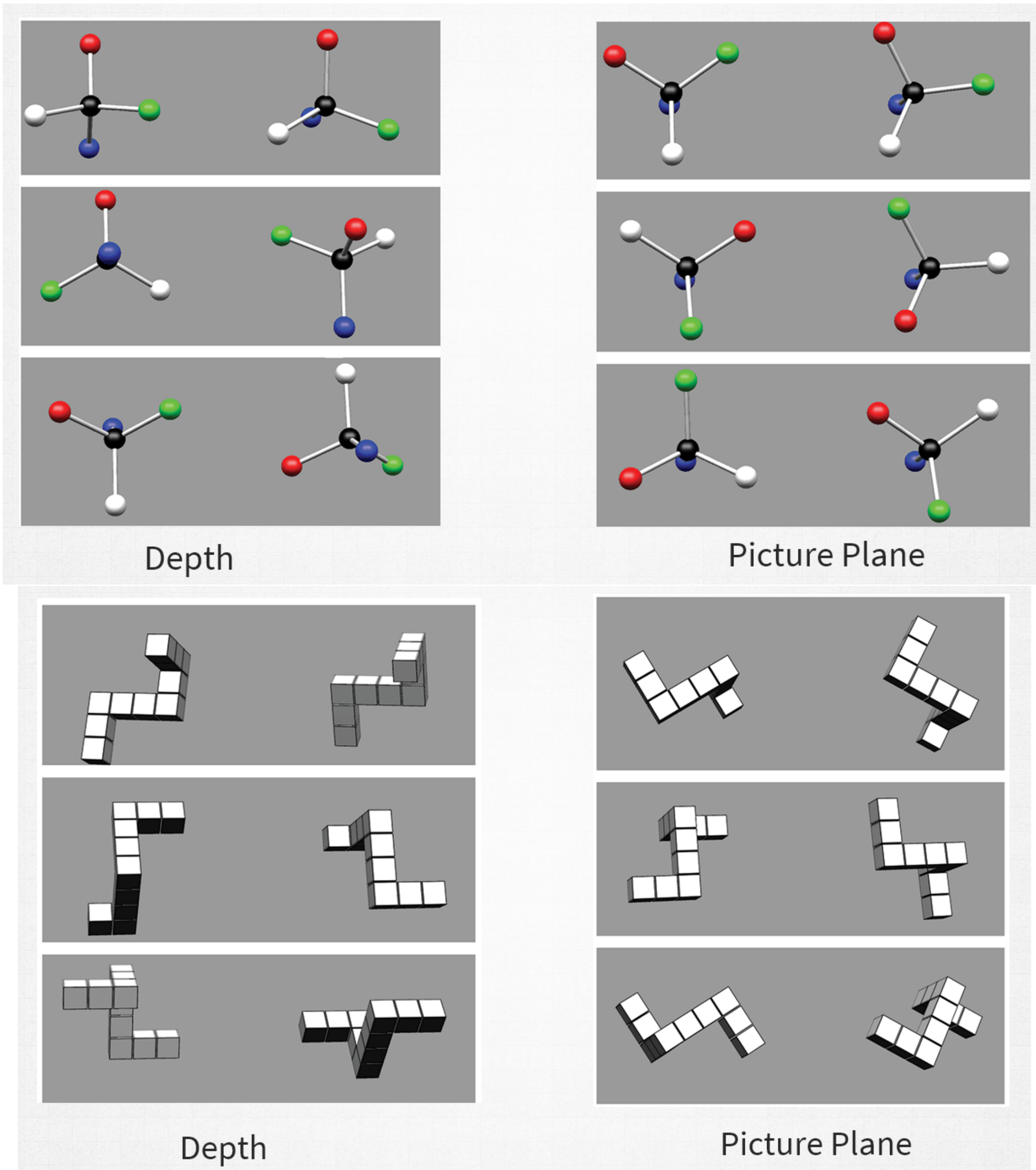

Figure 3.4: Example of twelve experimental trials. Tetrahedron and cube figure stimuli were rotated about depth and picture plane axes of rotation in separate trial blocks.

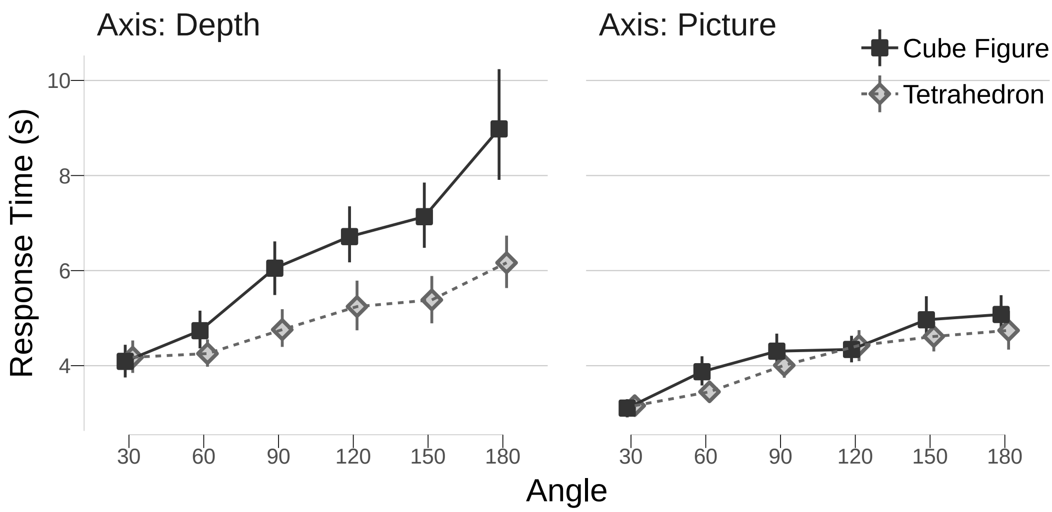

Figure 3.5: Response times as a function of angular disparity on 'same' (non-mirrored), correct trials. Within-subjects 95% CI are displayed.

Response time

The mean overall response time for the experiment was 4.78s and mean accuracy was 80.7%. Mean overall response times and within-subjects 95% confidence intervals for the different experimental conditions are shown in Figure 3.5. A 2 (Stimulus type: cube figure, tetrahedron) x 2 (Rotation axis: depth, picture) x 6 (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) ANOVA was conducted on raw response times for correct 'same' trials. Results for the ANOVA are summarized in Table 3.1. Angular disparity ($F[2.33, 58.32] = 38.44$, $MSE = 4.62$, $p < .001$, $\hat{\eta}^2_G = .147$), Stimulus ($F[1, 25] = 4.62$, $MSE = 17.60$, $p = .042$, $\hat{\eta}^2_G = .033$), and Axis ($F[1, 25] = 48.24$, $MSE = 7.30$, $p < .001$, $\hat{\eta}^2_G = .128$) affected response time. Significant two way interactions between Angle and Stimulus $F[2.6, 65.03] = 5.44$, $MSE = 2.13$, $p = .003$, $\hat{\eta}^2_G = .012$, Angle and Axis $F[3.3, 82.4] = 6.55$, $MSE = 1.78$, $p < .001$, $\hat{\eta}^2_G = .016$, Stimulus and Axis $F[1, 25] = 5.69$,

$MSE = 3.48$, $p = .025$, $\hat{\eta}^2_G = .008$ were observed, as well as the three-way interaction $F[2.89, 72.18] = 2.90$, $MSE = 2.07$, $p = .043$, $\hat{\eta}^2_G = .007$. This finding indicates that the cube figure stimuli had a significantly greater slope than the other conditions for depth-plane axis rotations. This finding suggests that a non-rotation

Table 3.1: Analysis of variance table for response times on 'same' (non-mirrored), correct trials.

Effect	F	df1	df2	MSE	p	ges
Angle	38.44	2.33	58.32	4.62	$< .001$	.147
Stimulus	4.62	1	25	17.60	.042	.033
Axis	48.24	1	25	7.30	$< .001$	.128
Angle × Stimulus	5.44	2.6	65.03	2.13	.003	.012
Angle × Axis	6.55	3.3	82.4	1.78	$< .001$	.016
Stimulus × Axis	5.69	1	25	3.48	.025	.008
Angle × Stimulus × Axis	2.90	2.89	72.18	2.07	.043	.007

Slope Analysis: Response Time

Slopes were calculated by computing separate simple linear regressions for each participant for each stimulus type and axis of rotation. For response time slopes, only identical correct trials were included. Each participant's data was subset by stimulus type and axis of rotation, and slopes were obtained by regressing response time over angle for each stimulus type and axis of rotation. Overall response time slopes are plotted in Figure 3.6. Response time slope was the dependent variable in a 2 (Stimulus type: cube figure, tetrahedron) x 2 (Rotation axis: depth, picture) ANOVA. Both main effects of Stimulus type ($F[1, 27] = 8.50$, $MSE = 0.00$, $p = .007$, $\hat{\eta}^2_G = .063$) and Axis of rotation ($F[1, 27] = 10.99$, $MSE = 0.00$, $p = .003$, $\hat{\eta}^2_G = .060$) reached significance. The main effects were qualified by the two-way interaction between Stimulus and Axis ($F[1, 27] = 6.92$, $MSE = 0.00$,

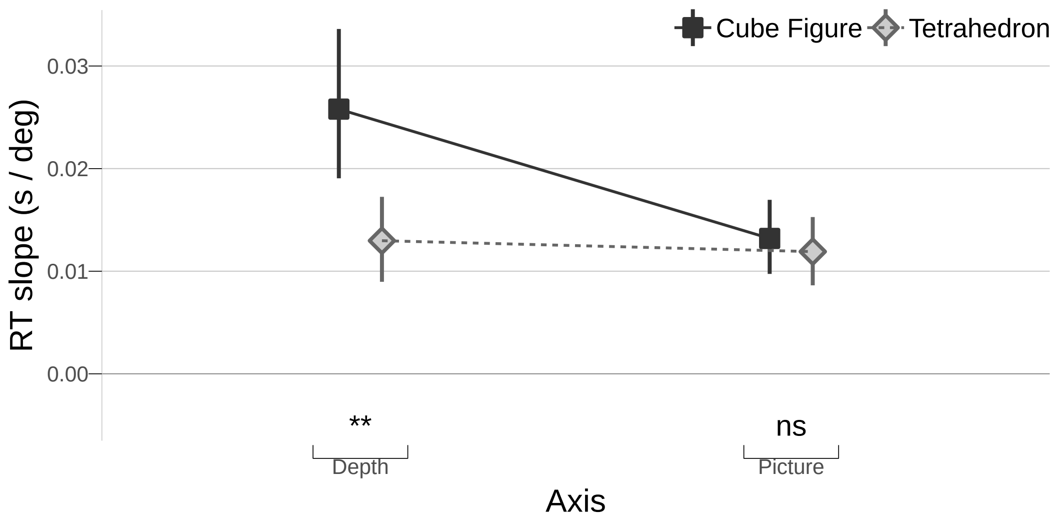

Figure 3.6: Response time slopes by axis of rotation and stimuli type.

$p = .014$, $\hat{\eta}^2_G = .043$). These results indicate that tetrahedron trials had significantly a smaller angular disparity effect on response time slopes than for cube figure trials.

Next, follow up analysis were conducted to understand how the effects interacted and how they compared to results from the classic mental rotation research. Marginal means were calculated for each condition. The slope coefficient unit is seconds per degree, so the reciprocal gives degrees / second. For the cube figure in depth plane, the marginal slope coefficient was 0.0258 sec/degree. This is equal to 38.8 degrees per second. For the cube figure in picture plane, the marginal slope coefficient was 0.0132 sec/deg or 75.8 degrees per second. Tetrahedron figures in depth plane was 77.1 degrees per second. Tetrahedron figure in picture plane was 84 degrees per second. These rates are comparable to the 60 degrees / second reported with the original Shepard and Metzler objects for depth rotations as described by Shepard and Cooper (1986). Importantly, depth-plane axis rotations of cube figures had significantly steeper slopes (i.e. greater angular disparity effect) than tetrahedrons while the slopes for the different stimuli types did not differ on picture plane rotations. This notable dissociation between angular disparity effects for the cube figures by axis of rotation offered evidence of a potential alternate strategy, thereby contaminating response times.

Table 3.2: Analysis of variance table for response time slopes on 'same', correct trials.

Effect	F	df1	df2	MSE	p	ges
Stimulus	8.50	1	27	0.00	.007	.063
Axis	10.99	1	27	0.00	.003	.060
Stimulus × Axis	6.92	1	27	0.00	.014	.043

Contrasts: Response time slope Results of pairwise contrasts between stimuli types are summarized in Table 3.3 for each axis of rotation. Contrasts between stimulus type revealed a significant difference between the slopes for the cube figure and tetrahedron stimuli on depth axis rotations ($\Delta M = 0.01$, 95% CI [0.01, 0.02], $t(53) = 3.93$, $p < .001$). However, for picture plane rotations there was no significant difference between stimuli types ($\Delta M = 0.00$, 95% CI [-0.01, 0.01], $t(53) = 0.39$, $p = .906$). There was

Table 3.3: Pairwise contrasts of response time slopes for 'same', trials.

	estimate	ci	statistic	p.value
Cube Figure - Tetrahedron, Depth	0.01	[0.01, 0.02]	3.93	$< .001$
Cube Figure - Tetrahedron, Picture	0.00	[-0.01, 0.01]	0.39	.906

no significant difference between cube figure and tetrahedron slopes for picture plane rotations. Most notably, the slope for depth axis tetrahedron trials was much lower than for the other three. Together, these results illustrate that response time was not as dependent on angular disparity for tetrahedron stimuli relative to cube figure stimuli.

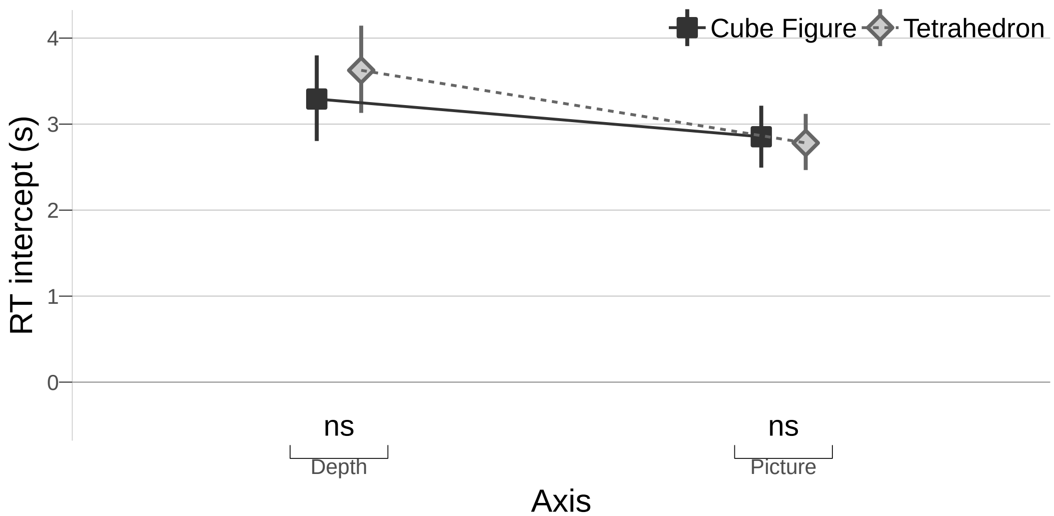

Figure 3.7: Response time intercepts by stimulus type and axis of rotation.

Intercept Analysis: Response time

Differences in strategy were expected to manifest as effects on slope– rather than on intercept, so analysis of intercepts was included as a matter of convention and for completeness. Figure 3.7 shows the results from the 2 (Stimulus type: cube figure, tetrahedron) x 2 (Rotation axis: depth, picture) ANOVA. As predicted, neither the main effects nor the two-way interaction reached significance. This can be interpreted to mean that both stimulus types required similar visual encoding time, and that differences in response time are due to the angular disparity effect.

Error rate

Overall error rates are depicted in Figure 3.8. A 2 (Stimulus type: cube figure, tetrahedron) x 2 (Rotation axis: depth, picture) x 6 (angular disparity: 30°, 60°, 90°, 120°,

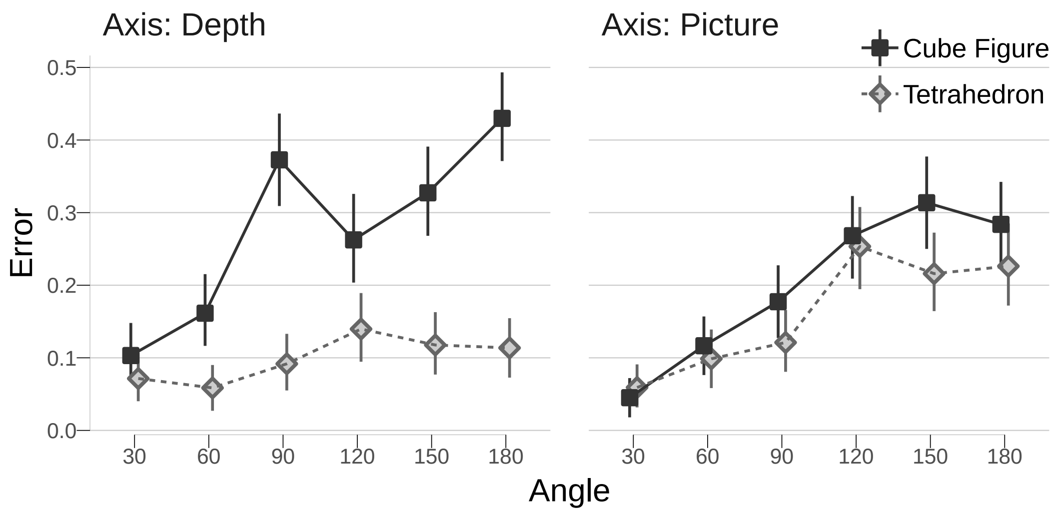

Figure 3.8: Error rate as a function of angular disparity by stimuli type and axis of rotation.

150°, 180°) ANOVA was conducted on error rates for "same" trials (see Table 3.4. Angular disparity ($F[3.21, 86.7] = 32.51$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .141$), Stimulus ($F[1, 27] = 14.68$, $MSE = 0.13$, $p = .001$, $\hat{\eta}^2_G = .085$), and Axis ($F[1, 27] = 0.15$, $MSE = 0.04$, $p = .698$, $\hat{\eta}^2_G = .000$) affected error rate. Significant two-way interactions between Angle x Stimulus ($F[3.45, 93.22] = 8.18$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .034$), Angle x Axis ($F[3.96, 106.81] = 3.90$, $MSE = 0.02$, $p = .006$, $\hat{\eta}^2_G = .017$), and Stimulus x Axis ($F[1, 27] = 25.18$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .037$) were observed, as well as the three-way interaction ($F[4.06, 109.71] = 2.47$, $MSE = 0.02$, $p = .048$, $\hat{\eta}^2_G = .011$). Analysis of error rates revealed a similarly flat slopes: the angular disparity effect on error rate was mostly absent for depth-plane rotations of tetrahedron stimuli. This finding offers additional evidence of an non-rotation cognitive process, and using this strategy leads to lower error rates across all angular disparities.

Table 3.4: Analysis of variance table for error rates on 'same' trials.

Effect	F	df1	df2	MSE	p	ges
Angle	32.51	3.21	86.7	0.03	$<.001$	.141
Stimulus	14.68	1	27	0.13	.001	.085
Axis	0.15	1	27	0.04	.698	.000
Angle × Stimulus	8.18	3.45	93.22	0.03	$<.001$	.034
Angle × Axis	3.90	3.96	106.81	0.02	.006	.017
Stimulus × Axis	25.18	1	27	0.03	$<.001$	.037
Angle × Stimulus × Axis	2.47	4.06	109.71	0.02	.048	.011

Slope Analysis: Error

Overall error rate slopes are shown in Figure 3.9. The slope coefficient obtained from the individual regressions on error rate was the dependent variable in a 2 (Stimulus type: cube figure, tetrahedron) x 2 (Rotation axis: depth, picture) ANOVA (see Table 3.5). A main effect of Stimulus type ($F[1,27] = 16.77$, $MSE = 0.02$, $p < .001$, $\hat{\eta}^2_G = .133$) was observed, while Axis of rotation ($F[1,27] = 3.09$, $MSE = 0.01$, $p = .090$, $\hat{\eta}^2_G = .016$) did not reach significance. The main effect of Stimulus was qualified by the two-way interaction between Stimulus and Axis ($F[1,27] = 4.97$, $MSE = 0.01$, $p = .034$, $\hat{\eta}^2_G = .034$). This finding indicates that tetrahedron trials had a significantly smaller angular disparity effect on response time slopes than for cube figure trials.

Table 3.5: Analysis of variance table for error rate slopes on 'same' (non-mirrored) trials.

Effect	F	df1	df2	MSE	p	ges
Stimulus	16.77	1	27	0.02	$<.001$	.133
Axis	3.09	1	27	0.01	.090	.016
Stimulus × Axis	4.97	1	27	0.01	.034	.034

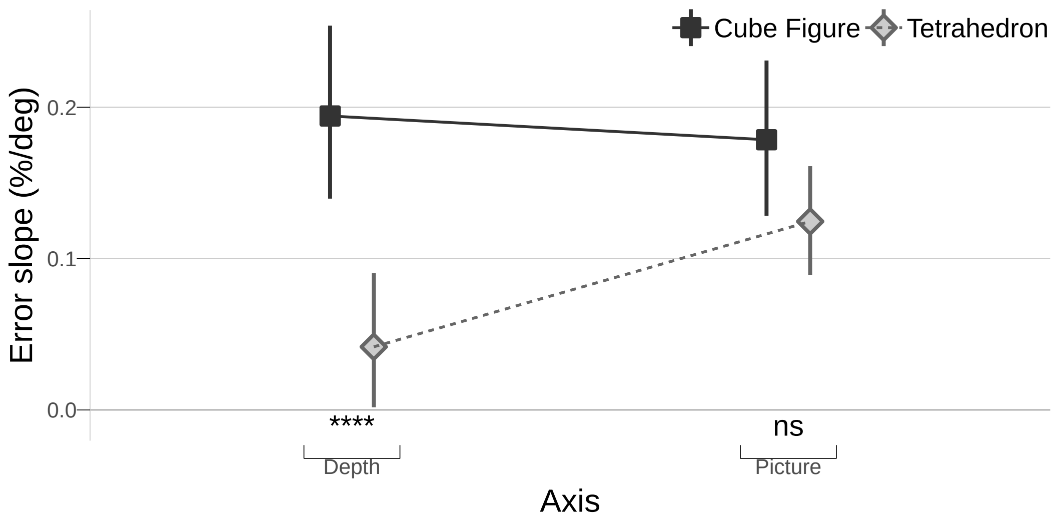

Figure 3.9: Error rate slopes by stimulus type and axis of rotation.

Contrasts: Error slope Pairwise contrasts on error rate slopes by stimulus are summarized in Table 3.6. Contrasts between stimulus type revealed a significant difference between the slopes of the cube figure and tetrahedron stimuli on depth axis rotations ($\Delta M = 0.01$, 95% CI [0.01, 0.02], $t(53) = 3.93$, $p < .001$). However, for picture plane rotations there was no significant difference between the slopes of the cube figure and tetrahedron stimuli ($\Delta M = 0.00$, 95% CI [-0.01, 0.01], $t(53) = 0.39$, $p = .906$). Overall,

Table 3.6: Pairwise contrast table by axis of rotation for error rate slopes for 'same' trials.

	estimate	ci	statistic	p.value
Cube Figure - Tetrahedron, Depth	0.15	[0.08, 0.23]	4.55	< .001
Cube Figure - Tetrahedron, Picture	0.05	[−0.02, 0.13]	1.61	.211

these results demonstrate that error rate was not as dependent on angular disparity for depth rotations of tetrahedron stimuli as the other conditions. Specifically, for depth plane rotations, the angular disparity effect on error rate was less pronounced on the tetrahedron stimuli compared to the cube figures. In other words, the effect of angular disparity on error rate was significantly greater for the cube figures on depth axis trials only. This finding provides further evidence that mental rotation may not be the only strategy utilized by participants for tetrahedron stimuli.

Paired slope correlations

If a common underlying cognitive process is operating for both stimuli types, then an individual's level of performance with one stimulus type should correlate with performance on the other stimulus type. If this is the case, then slopes should be positively correlated between stimuli within axis of rotation and individuals. The scatter plots in Figure 3.10 visualizes this relationship and shows that the correlation is only significant in the case of the picture plane rotation trials. For depth axis rotations, the correlation between stimuli on response time slope did not reach significance ($r = .18$, 95% CI $[-.20, .52]$, $t(26) = 0.96$, $p = .348$), however the correlation was significant for picture axis rotations ($r = .40$, 95% CI $[.03, .67]$, $t(26) = 2.24$, $p = .034$). Regarding error rate slope, neither correlation reached significance. The finding that slopes were only positively correlated on the picture plane axis trials and notably absent for the depth axis trials could suggest abstract reasoning being utilized depending on the axis of rotation. This dissociation provides another piece of converging evidence suggesting that an alternate non-rotation strategy is being used, at least by a portion of individuals for rotations of tetrahedrons about the depth-plane axis.

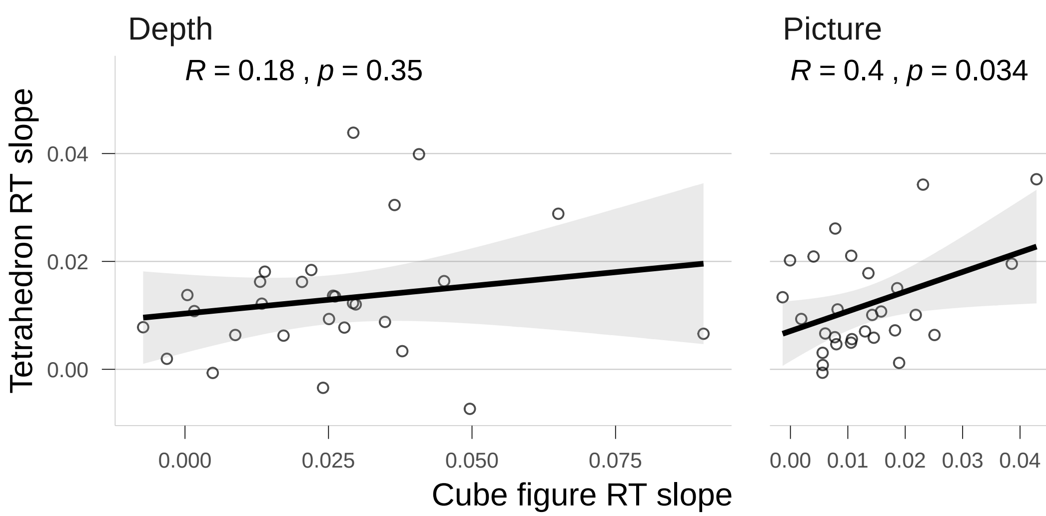

Figure 3.10: Correlations between individual slopes for each stimulus type by axis of rotation. Cube figure and tetrahedron response time slopes were positively correlated on picture plane trials only. There was no significant correlation between response time slopes for the different stimuli types on picture plane trials.

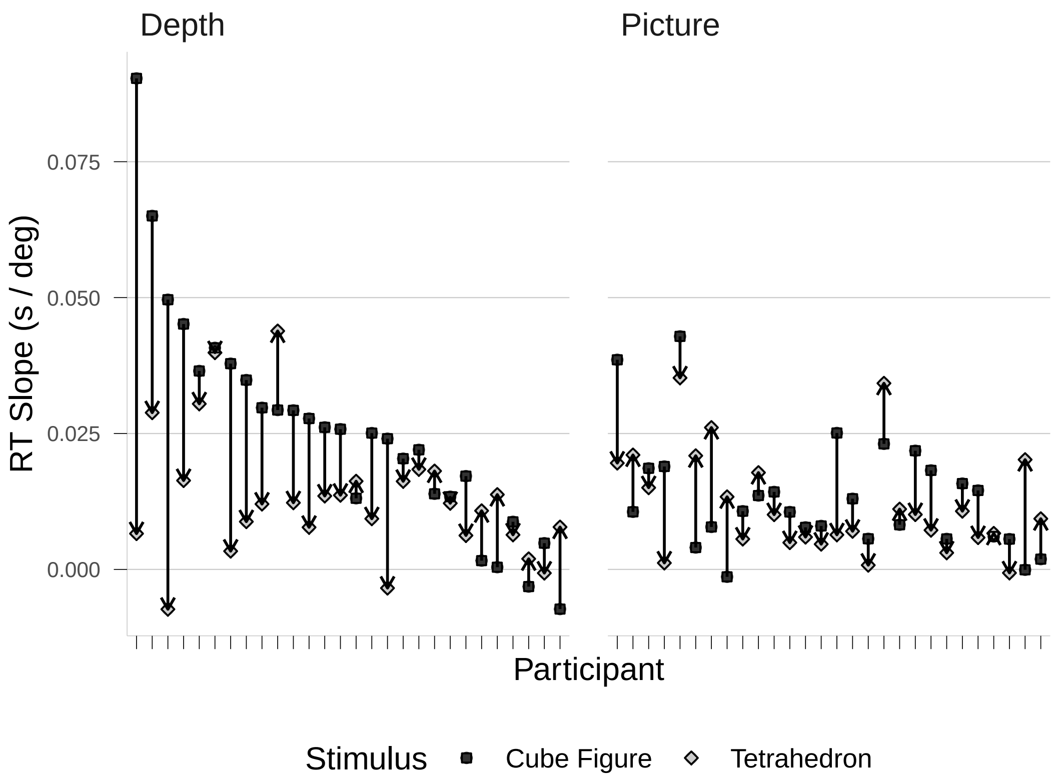

Figure 3.11: Individual paired slope coefficients for each stimulus by axis of rotation. A subset of participants had almost identical slope coefficients for both stimuli types, while others had disparate slopes. Individuals were ordered along the x-axis by overall slope for depth-plane rotations.

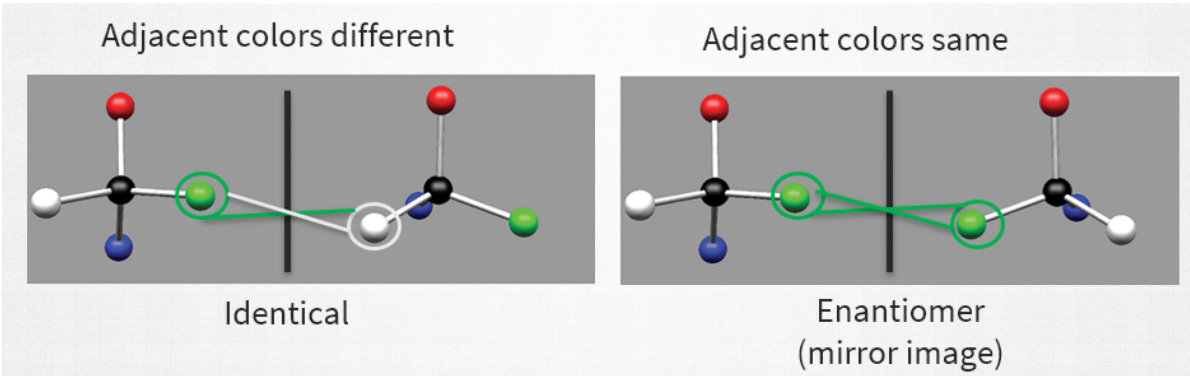

Figure 3.12: Specifically for trials with rotations about the horizontal depth-plane axis, the left and right elements do not move laterally across angular disparities. This allowed for the non-rotation strategy of comparing the relative locations of the adjacent elements, rather than mental rotation.

Individual Differences in slopes

Examination of participants' individual slopes offered further evidence of an alternate strategy being utilized by a subset of participants, particularly on the depth-axis trials. As seen in Figure 3.11, a subset of participants had a much lower slope on the tetrahedron stimuli compared to the cube figures for depth-plane rotations, but this was not true for picture-plane trials. Some participants had clearly separated slope coefficients depending on the stimulus type, while others approached both stimuli types using a mental rotation approach with similar slopes for both the cube figure and tetrahedron stimuli. Response patterns indicative of non-rotation strategy use were more apparent for the depth-axis trials.

An example solution using an alternate non-rotation strategy is shown in Figure 3.12. Use of the alternate strategy depends on whether people pick up that because all depth rotations are around the horizontal axis, the left and right elements (in this case the white and green balls) do not move laterally relative to each other across all angular disparities about the *pitch* principle axis. Interestingly, some participants picked up on the alternate strategy while others did not.

Experiment 2 focused on examining the development of alternate strategy use by providing additional alternating blocks of trials. Do participants pick up on the strategy rapidly, or does it take practice? Can we identify and predict users that utilize the non-rotation strategy?

Chapter 4

Experiment 2: Practice effects and strategy use

Introduction

The main purpose of this study to examine how alternate strategy use may change with practice during the task. In this experiment I focused specifically on depth-plane axis rotation trials. Trial presentation used pseudo-random list generation of 12 trials at each level of angular disparity. The aim here was to control for unwanted contextual influences that can introduce noise on repeated response measurement. Constraining the generation of serial trial order at this level eliminates the likelihood of receiving a long sequence of trials with similar angular disparities. This is important in this case for investigating strategy use over time, as a long sequence of easy trials at the beginning of the experiment may influence the adoption of one strategy over another. Participants completed six blocks of 48 trials (three per stimulus type, alternating and counterbalanced), or 288 trials total. In contrast to Experiment 1, here the experimental trials alternated between blocks of cube figure and tetrahedron trials. There were three blocks of trials for each stimulus, which allowed for examination of changes in performance across time.

Another purpose of this study was to examine whether adoption of alternate strategies was affected by using a stereo 3D display format. Since stereo display technologies have not traditionally been used in mental rotation experiments, it is important to determine how the display format impacts performance. Further, I wanted to rule out the possibility that the display format affected adoption of alternate strategies. Stereo and mono display formats were compared directly with a between-subjects design. If providing stereo viewing promotes use of alternate strategies, we should expect to see shallower slopes for individuals with the stereo display compared to individuals with the mono display. Observation of this effect was not expected, and I predicted that the additional depth cue of binocular disparity would not afford additional fidelity relevant for adoption of alternate strategies. Specifically, I predicted that performance would be sufficiently supported by monocular depth cues. On the other hand, it may be that providing stereo viewing aids in the process, of encoding the three-dimensional stimuli. If providing stereo viewing facilitates the encoding process then faster response times should be observed due a lower intercept compared to those in the mono display condition.

Method

Design.

Experiment 2 had a two (display dimensionality: stereo or mono) by two (stimulus: cube figure or tetrahedron) by three (Block: first, second, third) by six (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) mixed-design. Display dimensionality was varied between-subjects and the other factors were crossed within-subjects. The dependent measures were response time (on accurate *same* trials) error rate, as well as slope and intercept.

Participants.

Eighty-two college students from the psychology subject pool at a research university participated in the study in return for course credit. None of the participants had previously studied organic chemistry. All participants had normal or corrected-to-normal vision. Three participants were excluded due to experimenter error and 14 participants were excluded for near chance performance with less than 60% accuracy rates overall. Of the 65 participants that met the criteria for inclusion, 32 (13 female) were in the stereo condition and 33 (15 female) were in the mono display condition. The mean age of participants was 19.1 years (SD = 1.0, range = 18, 22).

Materials.

The materials were identical to Experiment 1 with modifications to the trial parameters. Trials consisted of depth axis rotations only, and the visual appearance of the stimuli were otherwise identical to Experiment 1. Trials were separated into three blocks of 48 trials for each stimulus type. Blocks of trials alternated between stimulus types to minimize ordering or carryover effects. Further, within each block of 48 trials, pseudo randomization was implemented on 12 trial sub-blocks containing both *same* and *mirror* trials for each of the six levels of angle once collapsed. The strategy questionnaire consisted of the following open-ended questions for each stimulus type: 1) “How did you go about solving these problems? For example, what tips would you give to someone who cannot solve the problems”, 2) “Did you use multiple strategies? If so, please describe how, when, and why you used each of them below", 3) "Did you learn or pick up on any new strategies or technique for solving the problems? Please describe". Two trained independent raters coded the strategy questionnaires to assign people as either rotators, alternate strategy users, or ambiguous. Inter-rater reliability was high (Cohen's Kappa $> .80$) and a third rater reconciled few disagreements.

Procedure.

Groups of up to 3 individuals were randomly assigned to a display type condition, either stereo or mono. The study was run in the same windowless laboratory room as Experiment 1. After providing informed consent, participants assigned to the stereo condition were instructed on the use of the 3D glasses and verified with the experimenter that the image displayed was fusing properly for them. Any adjustments were completed by the experimenter at this time. Next, instructions were read aloud by the experimenter while participants followed along with written instructions displayed on the monitor. Next, participants completed four practice trials with feedback, and were instructed to work as quickly and accurately as possible through the self-paced experimental trials. After the first block, participants completed four practice trials with the other stimulus type. There were no practice trials for all subsequent blocks. Finally, participants completed the self-report strategy questionnaire.

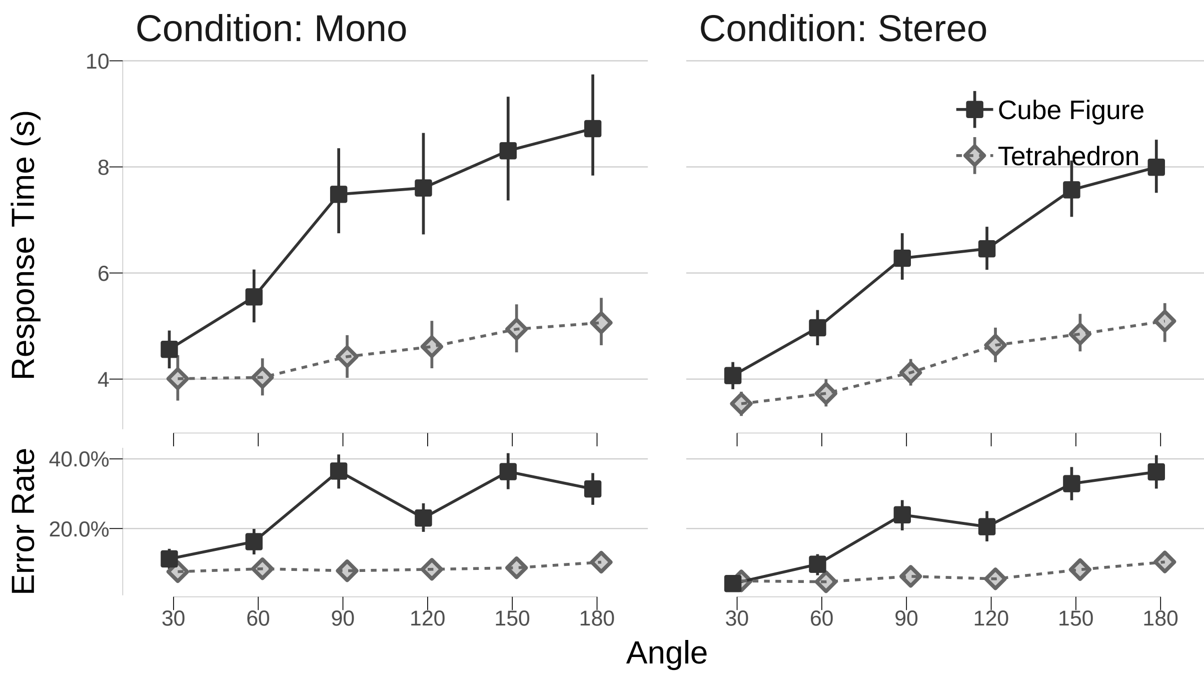

Figure 4.1: Response time and error rates as a function of angular disparity by display condition with 95% confidence intervals.

Results and Discussion

Data screening

Screening for fast and/or careless respondents showed that there were 14 participants that had less than .5s response times on more than 20% of total trials or near chance performance with less than 60% accuracy. Responses faster than .5s (0.37% of total trials) were not included in the analysis. There were no outlying long responses thanks to the pseudo-random blocked trial presentation.

Display dimensionality

Figure 4.1 shows the overall response times and error rates by angular disparity between the mono (left panels) and stereo display (right panels) conditions with 95% confidence intervals.

Response Time

The effect of display condition was examined using a 2 (Display condition: stereo, mono) x 2 (Stimulus type: cube figure, tetrahedron) x 6 (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) mixed design ANOVA on raw response times for correct *same* trials. Results of the ANOVA show that display condition did not have a significant main effect ($F[1, 62] = 1.09$, $MSE = 54.02$, $p = .300$, $\hat{\eta}^2_G = .008$) nor interactions with Angle ($F[2.45, 152.15] = 0.27$, $MSE = 7.01$, $p = .809$, $\hat{\eta}^2_G = .001$), Stimulus ($F[1, 62] = 1.23$, $MSE = 25.51$, $p = .271$, $\hat{\eta}^2_G = .005$), or Angle x Stimulus ($F[2.76, 171.08] = 0.91$, $MSE = 5.48$, $p = .433$, $\hat{\eta}^2_G = .002$) on overall response time. This finding suggests that the two display types did not impact response time for either stimulus type.

Error Rate

The effect of display condition was examined using a 2 (Display condition: stereo, mono) x 2 (Stimulus type: cube figure, tetrahedron) x 6 (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) mixed design ANOVA on error rates for *same* trials. Results of the ANOVA show that display condition did not have a significant main effect ($F[1, 62] = 1.23$, $MSE = 0.11$, $p = .271$, $\hat{\eta}^2_G = .010$) nor interactions with Angle ($F[4.27, 264.77] = 1.04$, $MSE = 0.01$, $p = .389$, $\hat{\eta}^2_G = .002$), Stimulus ($F[1, 62] = 0.30$, $MSE = 0.04$, $p = .586$, $\hat{\eta}^2_G = .001$), or Angle x Stimulus ($F[4.43, 274.62] = 0.86$, $MSE = 0.01$, $p = .498$, $\hat{\eta}^2_G =$

.002) on overall error rate. This finding suggests that display type did not affect error rates.

RT Intercept

The effect of display condition was examined using a 2 (Display condition: stereo, mono) x 2 (Stimulus type: cube figure, tetrahedron) x 6 (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) mixed design ANOVA on error rates for *same* trials. Results of the ANOVA show only a main effect on Block ($F[1.68, 104.24] = 38.84$, $MSE = 5.06$, $p < .001$, $\hat{\eta}^2_G = .142$). No significant effect of Display condition ($F[1, 62] = 1.68$, $MSE = 10.65$, $p = .199$, $\hat{\eta}^2_G = .009$), nor Stimulus ($F[1, 62] = 0.10$, $MSE = 7.73$, $p = .750$, $\hat{\eta}^2_G = .000$), nor Angle x Stimulus ($F[1.63, 101.18] = 0.95$, $MSE = 3.26$, $p = .374$, $\hat{\eta}^2_G = .003$) on intercept were observed. As shown in Figure 4.2, response time intercepts for the stereo and mono display condition did not differ, and thus does not support the hypothesis that added display dimensionality aid encoding to produce smaller intercepts.

RT Slope

The effect of display type on response time slope was examined using a 2 (Display type: stereo, mono) x 2 (Stimulus type: cube figure, tetrahedron) x 6 (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) mixed design ANOVA for correct *same* trials. The results showed that display type only had a main effect on Block ($F[1.71, 106.01] = 3.51$, $MSE = 0.00$, $p = .040$, $\hat{\eta}^2_G = .013$). No significant effect of Display type ($F[1, 62] = 0.03$, $MSE = 0.00$, $p = .868$, $\hat{\eta}^2_G = .000$), nor Stimulus ($F[1, 62] = 1.43$, $MSE = 0.00$, $p = .237$, $\hat{\eta}^2_G = .006$), nor Angle x Stimulus ($F[1.71, 106.06] = 0.70$, $MSE = 0.00$, $p = .479$, $\hat{\eta}^2_G = .002$) on slope were observed. As shown in Figure 4.3, response time slopes for the stereo and mono display condition did not differ, this finding suggests that angular

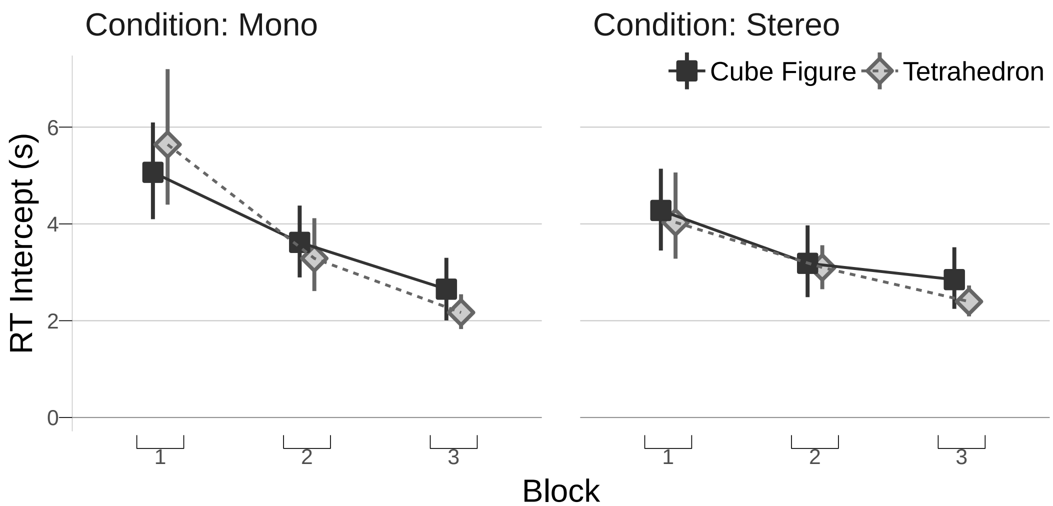

Figure 4.2: Response time intercepts as a function of trial block by stimulus type.

disparity had similar effects on response time slopes across display types. Overall, the two display types produced similar performance for both the tetrahedron and cube figure stimuli. Display condition was not predicted to have a significant impact on performance, as the addition of binocular depth cues should not afford additional strategies versus the mono display format. Display dimensionality was examined to rule out its impact on response times or error rates, so display type collapsed in the following analyses to ease interpretation.

Practice Effects

Response Time

Figure 4.4 shows overall response times across blocks of trials for each stimuli type. A 2 (Stimulus type: cube figure, tetrahedron) x 3 (trial Block: first, second, third) x 6 (angu-

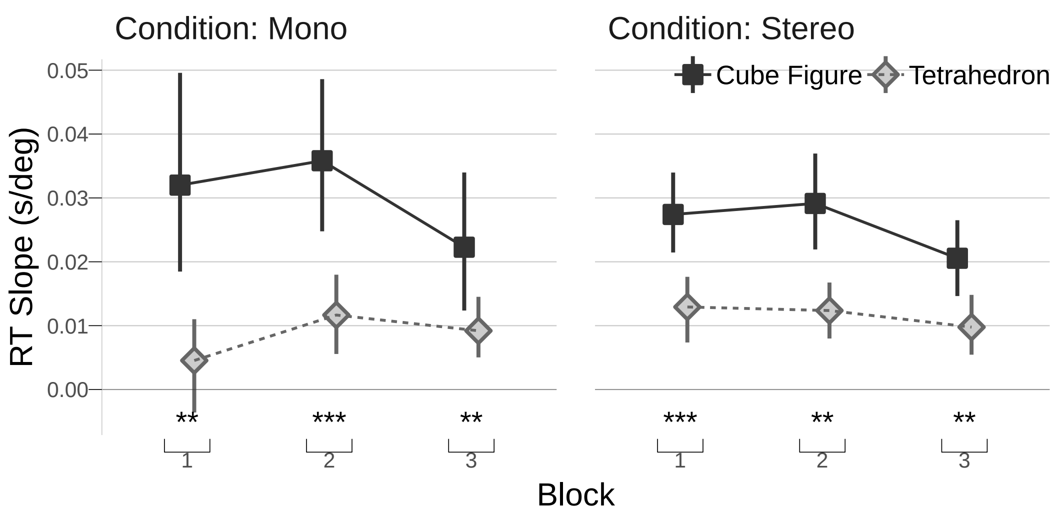

Figure 4.3: Response time slope as a function of trial block by display condition.

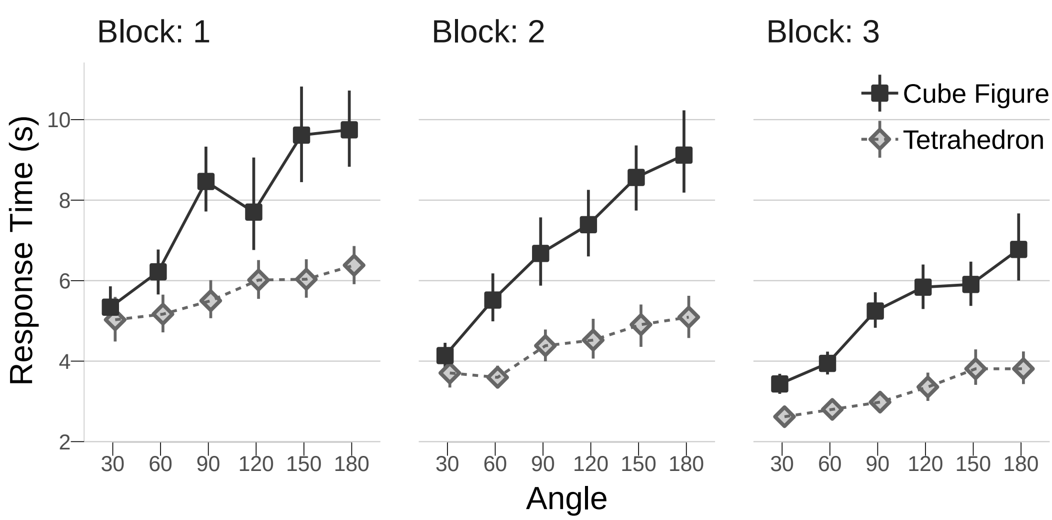

Figure 4.4: Response times as a function of angular disparity by trial block on ‘same’, correct trials.

Table 4.1: Analysis of variance table for response times on 'same' (non-mirrored), correct trials.

Effect	F	df1	df2	MSE	p	ges
Angle	54.99	2.58	100.54	10.10	$< .001$	.104
Stimulus	41.30	1	39	28.18	$< .001$	.086
Block	78.77	1.55	60.31	11.10	$< .001$	.099
Angle × Stimulus	12.67	2.81	109.61	9.04	$< .001$	.025
Angle × Block	1.57	4.4	171.59	9.45	.178	.005
Stimulus × Block	1.32	1.47	57.28	12.25	.268	.002
Angle × Stimulus × Block	1.72	5.66	220.85	6.37	.122	.005

lar disparity: 30°, 60°, 90°, 120°, 150°, 180°) within-design ANOVA conducted on response times for correct *same* trials (see Table 4.1). Angular disparity ($F[2.58, 100.54] = 54.99$, $MSE = 10.10$, $p < .001$, $\hat{\eta}^2_G = .104$), Stimulus ($F[1, 39] = 41.30$, $MSE = 28.18$, $p < .001$, $\hat{\eta}^2_G = .086$), and Block ($F[1.55, 60.31] = 78.77$, $MSE = 11.10$, $p < .001$, $\hat{\eta}^2_G = .099$) affected response time. A significant two-way interaction between Stimulus and Block ($F[2.81, 109.61] = 12.67$, $MSE = 9.04$, $p < .001$, $\hat{\eta}^2_G = .025$) was observed. In general, this replicated the finding from Experiment 1 that response time was not as dependent on angular disparity for tetrahedron stimuli versus the cube figure stimuli.

Response Time Slope Figure 4.5 shows overall response time slopes across blocks of trials for each stimulus type. Smaller slopes were observed for the tetrahedron stimuli trials, and this effect was persistent across trial blocks. Response time slope was the dependent variable in a 2 (Stimulus type: cube figure, tetrahedron) x 3 (trial Block: first, second, third) ANOVA. Both main effects of Stimulus type ($F[1, 63] = 31.50$, $MSE = 0.00$, $p < .001$, $\hat{\eta}^2_G = .120$) and trial Block ($F[1.72, 108.24] = 3.54$, $MSE = 0.00$, $p = .039$, $\hat{\eta}^2_G = .013$) reached significance. No significant interaction was found for Stimulus and Block ($F[1.7, 107.36] = 2.54$, $MSE = 0.00$, $p = .092$, $\hat{\eta}^2_G = .007$). The response time

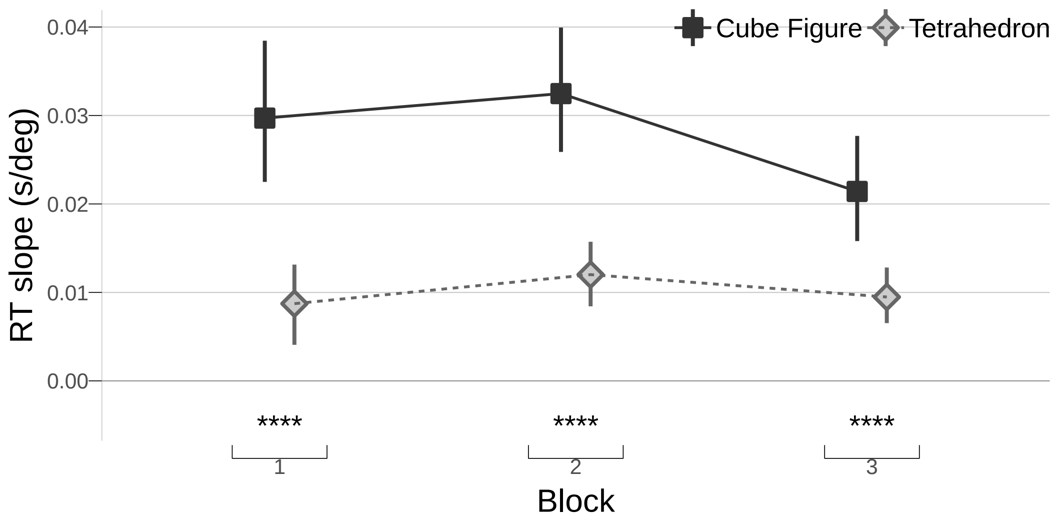

Figure 4.5: Response time slopes as a function of trial block by stimulus type.

slope contrasts revealed that the angular disparity effect was greater on cube figure trials than for tetrahedron trials, and this effect was persistent across trial blocks (all $p <$ .015). In general, the response time slopes were larger for cube figures on all blocks of trials– indicating that alternate strategy use developed during the first trial block with tetrahedrons.

Response time intercept Figure 4.6 shows overall response time intercepts across blocks of trials for each stimulus type. Similar response time intercepts were observed across all three blocks of trials. A 2 (Stimulus type: cube figure, tetrahedron) x 3 (trial Block: first, second, third) ANOVA was conducted on the response time intercept. The effect of trial Block ($F[1.64, 103.35] = 37.17$, $MSE = 5.42$, $p < .001$, $\hat{\eta}^2_G = .139$) reached significance, while there was no significant effect of Stimulus type ($F[1, 63] = 0.37$, $MSE = 7.62$, $p = .546$, $\hat{\eta}^2_G = .001$). Both stimuli shared similar response time intercepts, and the

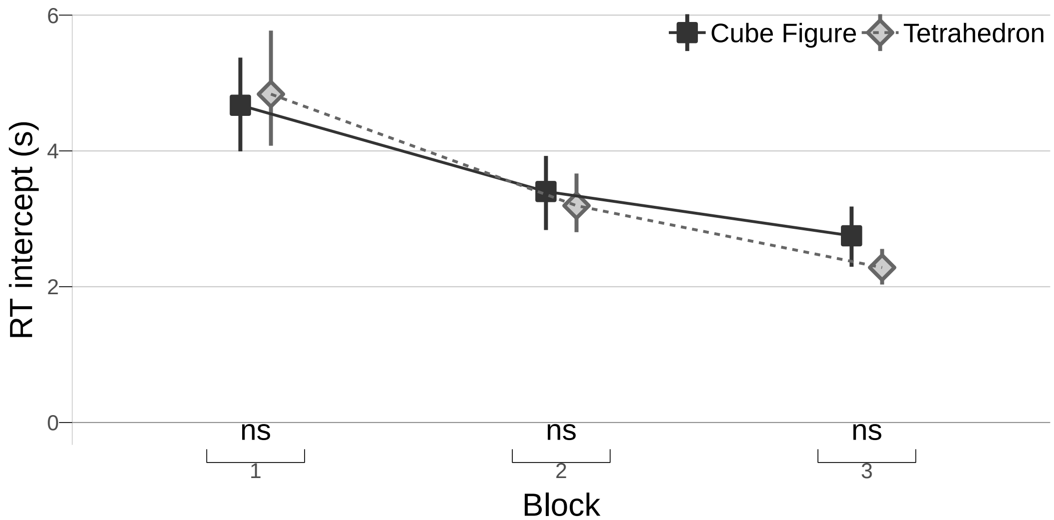

Figure 4.6: Response time intercepts as a function of trial block by stimulus type.

intercepts for each stimulus decreased about the same amount with practice. This can be interpreted to mean that the two stimulus types required similar visual encoding time, and that differences in response time are generally due to the angular disparity effect – the mental rotation signature.

Error rate

Figure 4.7 shows overall error rates across blocks of trials for each stimulus type. Smaller error rate slopes were observed for all three blocks of the tetrahedron stimuli trials. A 2 (Stimulus type: cube figure, tetrahedron) x 3 (trial Block: first, second, third) within-design ANOVA was conducted on error rates for *same* trials. Angular disparity ($F[3.63, 228.41] = 41.30$, $MSE = 0.05$, $p < .001$, $\hat{\eta}^2_G = .072$) and Stimulus ($F[1, 63] = 116.89$, $MSE = 0.13$, $p < .001$, $\hat{\eta}^2_G = .130$) had main effects on error rate. Significant interactions between Angle x Stimulus ($F[4.08, 257.06] = 25.07$, $MSE = 0.05$, $p < .001$,

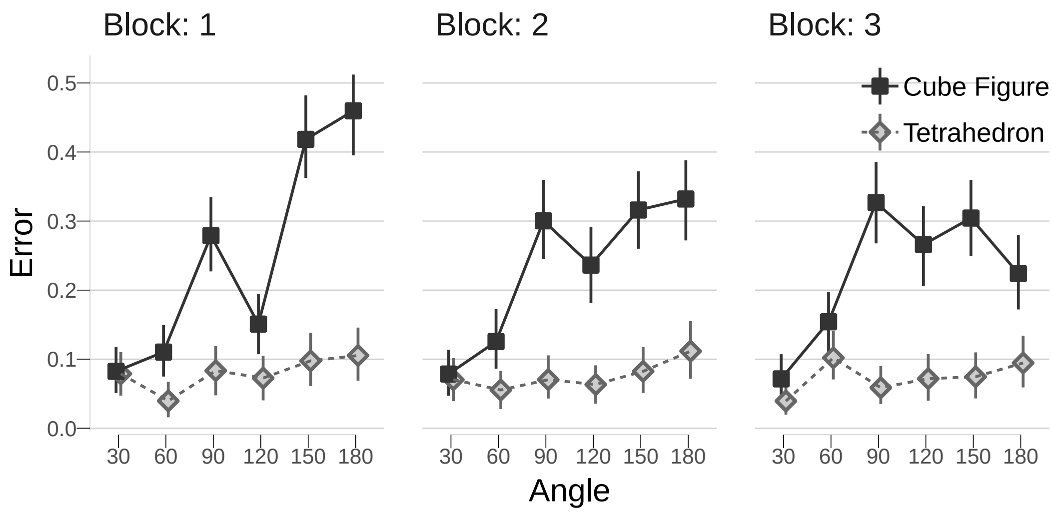

Figure 4.7: Error rates as a function of angular disparity by stimulus type and trial block.

$\hat{\eta}^2_G = .048$), Angle x Block ($F[7.88, 496.14] = 5.58$, $MSE = 0.04$, $p < .001$, $\hat{\eta}^2_G = .016$), and Angle x Stimulus x Block ($F[7.45, 469.12] = 4.52$, $MSE = 0.04$, $p < .001$, $\hat{\eta}^2_G = .013$) were observed. See Table 4.2.

Table 4.2: Analysis of variance table for error rates on 'same' trials.

Effect	F	df1	df2	MSE	p	ges
Angle	41.30	3.63	228.41	0.05	< .001	.072
Stimulus	116.89	1	63	0.13	< .001	.130
Block	0.98	1.9	119.98	0.05	.376	.001
Angle × Stimulus	25.07	4.08	257.06	0.05	< .001	.048
Angle × Block	5.58	7.88	496.14	0.04	< .001	.016
Stimulus × Block	0.40	1.72	108.1	0.05	.639	.000
Angle × Stimulus × Block	4.52	7.45	469.12	0.04	< .001	.013

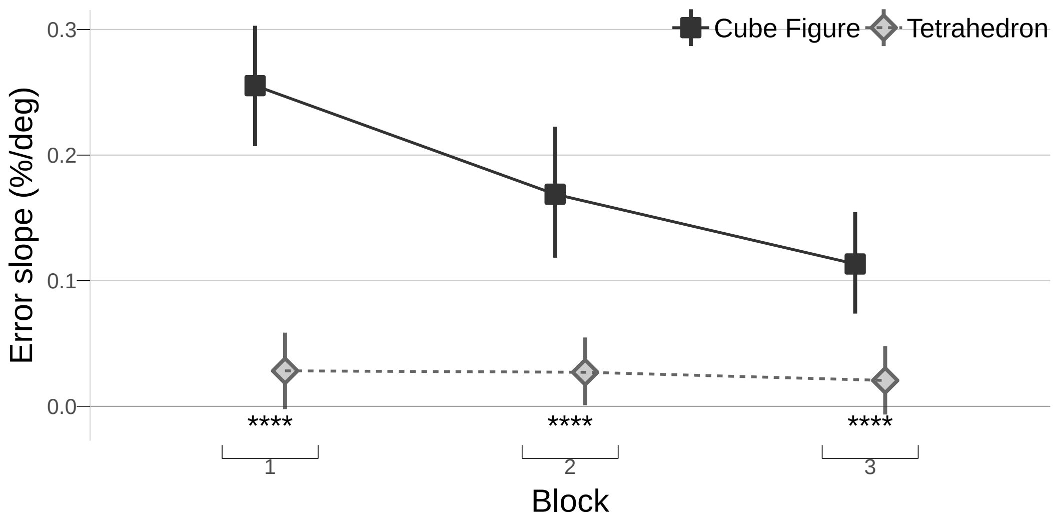

Figure 4.8: Error rate slopes as a function of trial block and stimulus type.

Error rate slope Figure 4.8 shows overall error rate slopes across blocks of trials for each stimulus type. Shallower error rate slopes were observed on all three blocks of the tetrahedron stimuli trials, and this effect decreased with practice. Larger slope values indicated that people made more errors on trials with larger angular disparities only on the cube figure trials. A 2 (Stimulus type: cube figure, tetrahedron) x 3

Table 4.3: Analysis of variance table for error rate slopes on 'same' (non-mirrored) trials.

Effect	F	df1	df2	MSE	p	ges
Stimulus	51.49	1	63	0.04	$< .001$	.186
Block	10.29	1.97	124.28	0.02	$< .001$	.035
Stimulus $\times$ Block	8.72	1.91	120.41	0.02	$< .001$	.029

(trial Block: first, second, third) within-design ANOVA conducted on error rate slopes for *same* trials (see Table 4.3). Both main effects of Stimulus type ($F[1, 63] = 51.49$,

$MSE = 0.04$, $p < .001$, $\hat{\eta}^2_G = .186$) and trial Block ($F[1.97, 124.28] = 10.29$, $MSE = 0.02$, $p < .001$, $\hat{\eta}^2_G = .035$) reached significance. The main effects were qualified by the two-way interaction between Stimulus and Block ($F[1.91, 120.41] = 8.72$, $MSE = 0.02$, $p < .001$, $\hat{\eta}^2_G = .029$). Overall, shallower error rate slopes were observed on the tetrahedron stimuli than the cube figure stimuli. The interaction between Stimulus and Block suggested that the angular disparity effect on error rate decreased with practice only for cube figures. In other words, the angular disparity effect on error rate slope was significantly smaller for tetrahedrons than for cube figures and this effect diminished with practice.

Strategy use

Two trained independent raters coded the strategy questionnaires to assign people as either rotators, alternate strategy users, or ambiguous. Inter-rater reliability was high (Cohen's Kappa $> .80$) and a third rater reconciled few disagreements, and three ambiguous cases were not included. Table 4.4 shows the overall mean performance of each strategy group based on the self-reported strategy questionnaire. In general, those who reported using the analytic strategy were faster and more accurate with the tetrahedron trials than those who did not report using the alternate strategy (rotators).

Table 4.4: Overall perfomance by self-reported strategy.

Self-report	Stimulus	RT (s)	slope (s/deg)	Error rate	n
Alternate	Cube Figure	7.68	0.03	0.3	34
Alternate	Tetrahedron	4.96	0.009	0.06	34
Rotate	Cube Figure	8.32	0.03	0.3	26
Rotate	Tetrahedron	7.04	0.006	0.1	26

Model-Based Clustering Analysis

I conducted a model-based cluster analysis to classify participants based solely on observable behavior. The goal here was to classify participants by strategy based on their response time slopes. I focused on response time slopes for the third block of tetrahedron trials, as alternate strategies were more likely to develop with practice. Response time slopes for the last block of tetrahedron trials for each participant were analyzed using Mclust (Scrucca, Fop, Murphy, & Raftery, 2016). An advantage of model-based clustering is that it is unsupervised, so there was no decision making regarding the type or number of models to fit by the researcher. This objectivity is desirable when the experimental task may not be cognitively penetrable to the participants, as in the present study. Acknowledging the limitations of self-reported cognitive strategy use– an estimate of the predictive accuracy of the resulting model was obtained using labeled data.

If multiple strategies are being used and one is a non-rotation strategy, then one group of people should have a distribution of slopes that is both smaller and less variant than others. A cluster of positive slopes could be interpreted as participants reliably having an angular disparity effect– the signature of a mental rotation strategy. In other words, response time slopes should in general be shallower for participants who use an alternate strategy versus those who use a mental rotation strategy. Figure 4.9 shows the histogram of RT slopes with the estimated density estimate curve (upper left). There was a large concentration of slopes approaching zero as well as a wide range of positive slopes. Two classes of models were evaluated: equal variance and unequal variance models, each fit with different numbers of mixture components. The lower left panel of Figure 4.9 plots the BIC model selection criteria and shows two-component variable-variance model produced the best fit indicated by the BIC statistic. The estimated versus empirical cumulative density function plot (lower center) and Q-Q plot (lower right) depict good

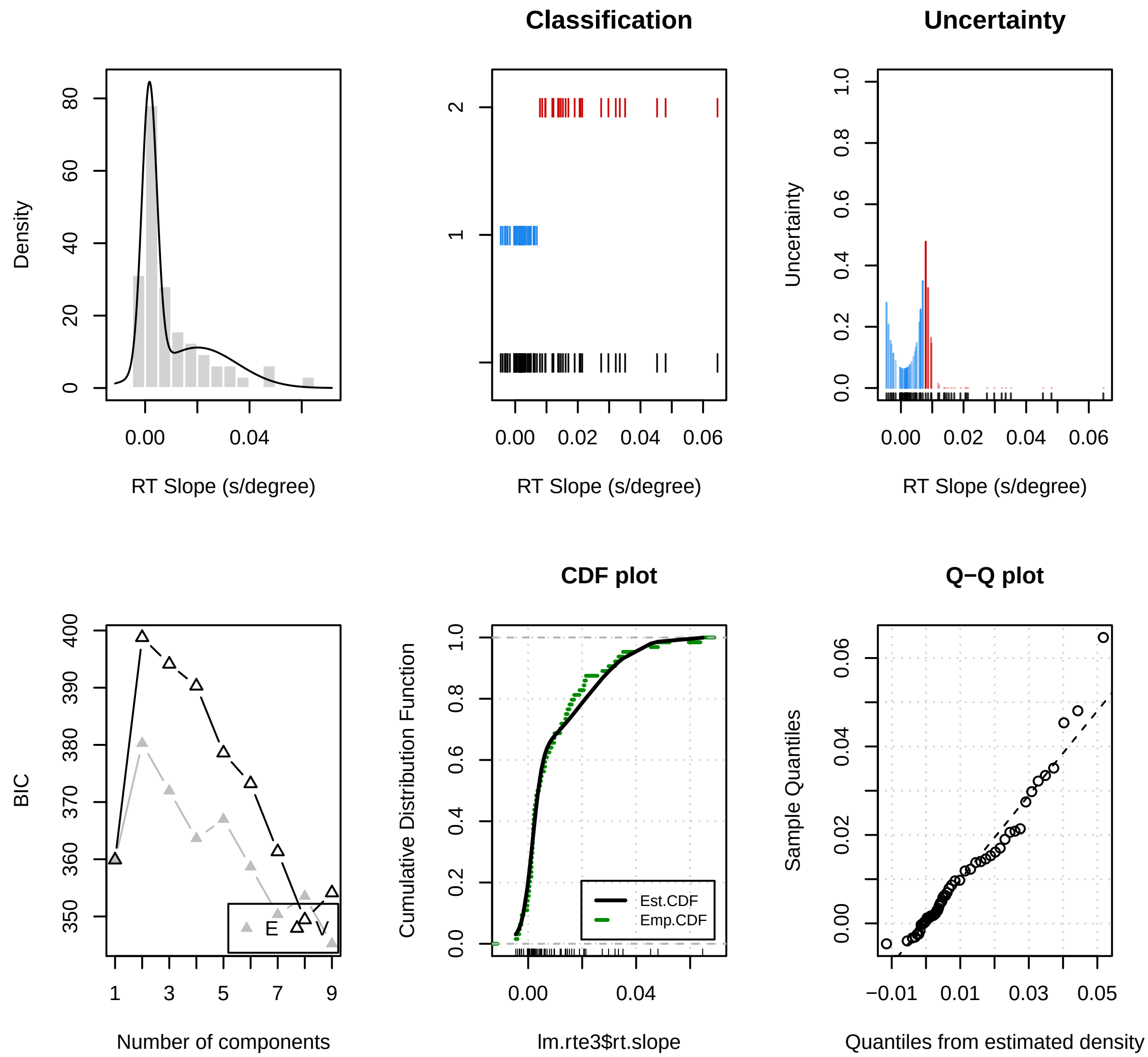

Figure 4.9: Estimated density plot (top left) and model-based clustering analysis of response time slopes. The preferred variable-variance two-component Gaussian finite mixture model had the highest Bayesian Information Criterion (bottom left).

fit diagnostics. The resulting classification table is visualized (upper center) along with the uncertainty associated with everyone's predicted class (upper right).

The summary of the fitted uni-variate, unequal variance Gaussian finite mixture model by the Expectation Maximization algorithm produced a clustering table with 40 participants in cluster 1 and 24 in cluster 2. Cluster 1 had response time slopes with both mean and variance (M = 611.14 deg/s, S = 8.3814e-06 s/deg) an order of a magnitude smaller than cluster 2 (M = 49.71 deg/s, S = 0.0002296 s/deg). Parametric bootstrap likelihood ratio test was used to determine the optimal number of mixture components for the uni-variate unequal variance model. Findings showed that the single component model could be rejected in favor of the two-component model according to the likelihood ratio test score (LRTS) p-value of .001. The two-component model however could not be rejected in favor of the 3 component model (LRTS = 7.82, $p > .05$).

The preferred model aligned well with the prediction that a portion of participants (in cluster 1) may pick up on an orientation independent strategy while others (in cluster 2) continue using mental rotation. The uncertainty of each participants' class assignment is visualized in the upper-right panel of Figure 4.9, with larger uncertainty for classifications around boundary cases.

Supervised classification To evaluate the preferred model, I determined its predictive accuracy on labeled training data. Discriminant analysis was conducted to obtain the predictive error of the classifier. Results of the discriminant analysis are visualized in Figure 4.10. The model had a predictive error of 26% on the training data set. This could be interpreted as the classifier agreeing with the self-reported strategy 74% of the time. This seemed reasonable given the known unreliability of self-reports of cognitive strategy, especially in cases participants switch or use multiple strategies. For example,

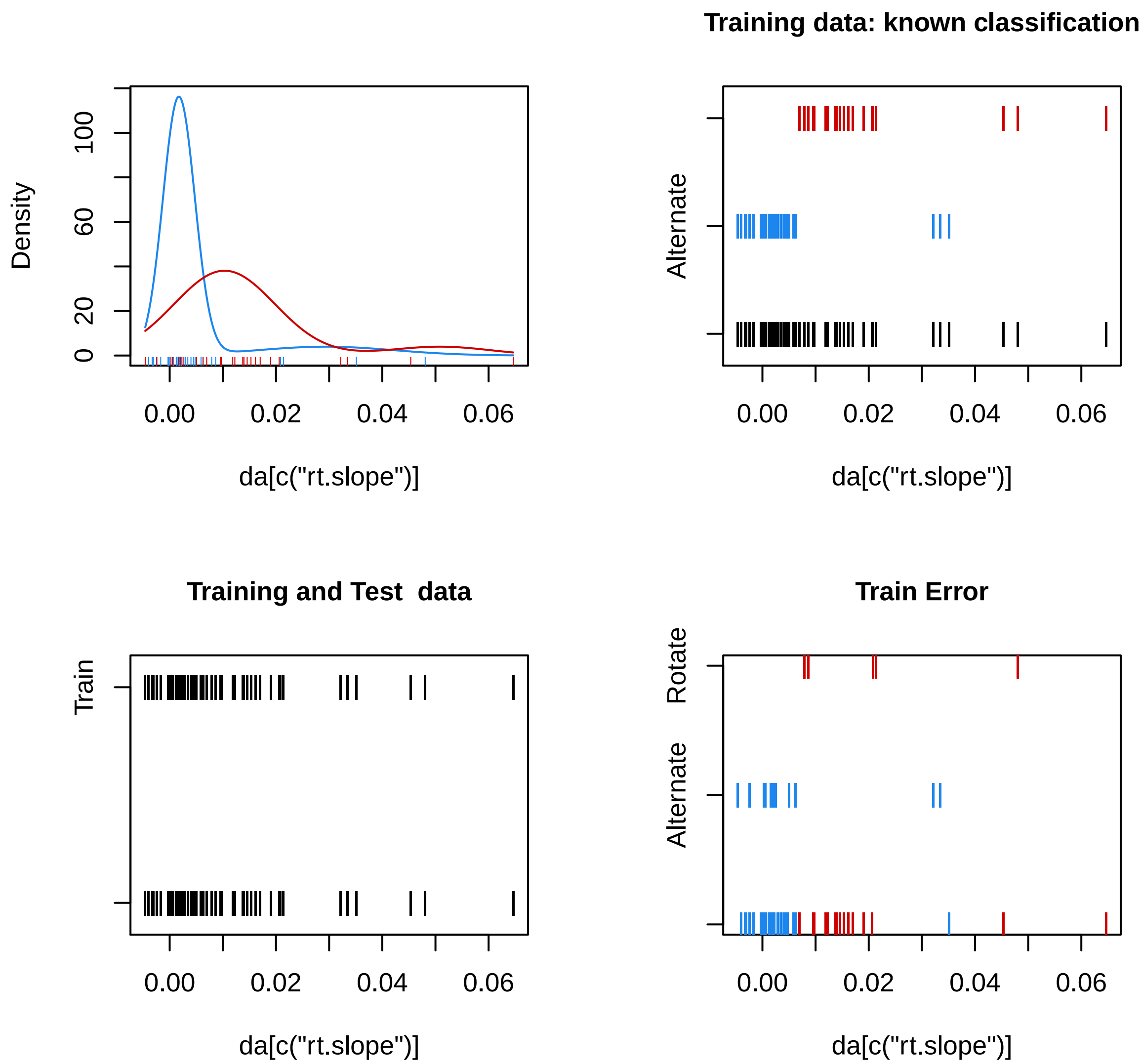

Figure 4.10: Discriminant analysis results summary. Top-left panel shows the two estimated distributions of response time slopes based on self-report. Known classifications (self-reported strategy) are displayed top-right. Bottom-right panel displays where the training errors occurred. The mixture-based discriminant analysis had a predictive error of 26%.

Table 4.5: Clustering results table: classifier by self-report

Self.report	cluster_1	cluster_2
Alternate	29	5
Rotation	11	15

participants may not report their initial strategy they used or vice-verse. Self-reported versus cluster analysis derived strategy classifications are shown in Table 4.5. K-fold cross validation was conducted to validate the predictive accuracy of the model. Here, the original data is randomly partitioned into k equal size subsets (*folds*). One of the k folds is held back as the validation data for model testing and the remaining k-1 for training; and this is repeated k times. Each of the k estimations of predictive error are subsequently averaged to reduce variance. 10-folds were randomly shuffled, and the predictive error produced was $\mathit{Error} = 30\%$, $\mathit{SE} = 6.64\%$.

How did strategy use impact performance?

Strategy use on response time Figure 4.11 shows overall response times for the two different strategy groups. Participants who picked up on the analytic strategy had shallower slopes for the tetrahedron figures, and these slopes became flatter across blocks of trials. Participants classified as rotators had similar angular disparity effects for both the tetrahedron and cube figure stimuli.

Strategy use on response time slope Both main effects of Stimulus type ($F[1, 62] = 28.61$, $\mathit{MSE} = 0.00$, $p < .001$, $\hat{\eta}^2_G = .118$) and strategy Classification ($F[1, 62] = 7.43$, $\mathit{MSE} = 0.00$, $p = .008$, $\hat{\eta}^2_G = .033$) reached significance. The main effects were qualified by the two-way interaction between Block and strategy Classification ($F[1.78, 110.22] = 6.76$, $\mathit{MSE} = 0.00$, $p = .003$, $\hat{\eta}^2_G = .024$). In general, participants classified as using an alternate strategy had tetrahedron slopes that approached zero during the third block

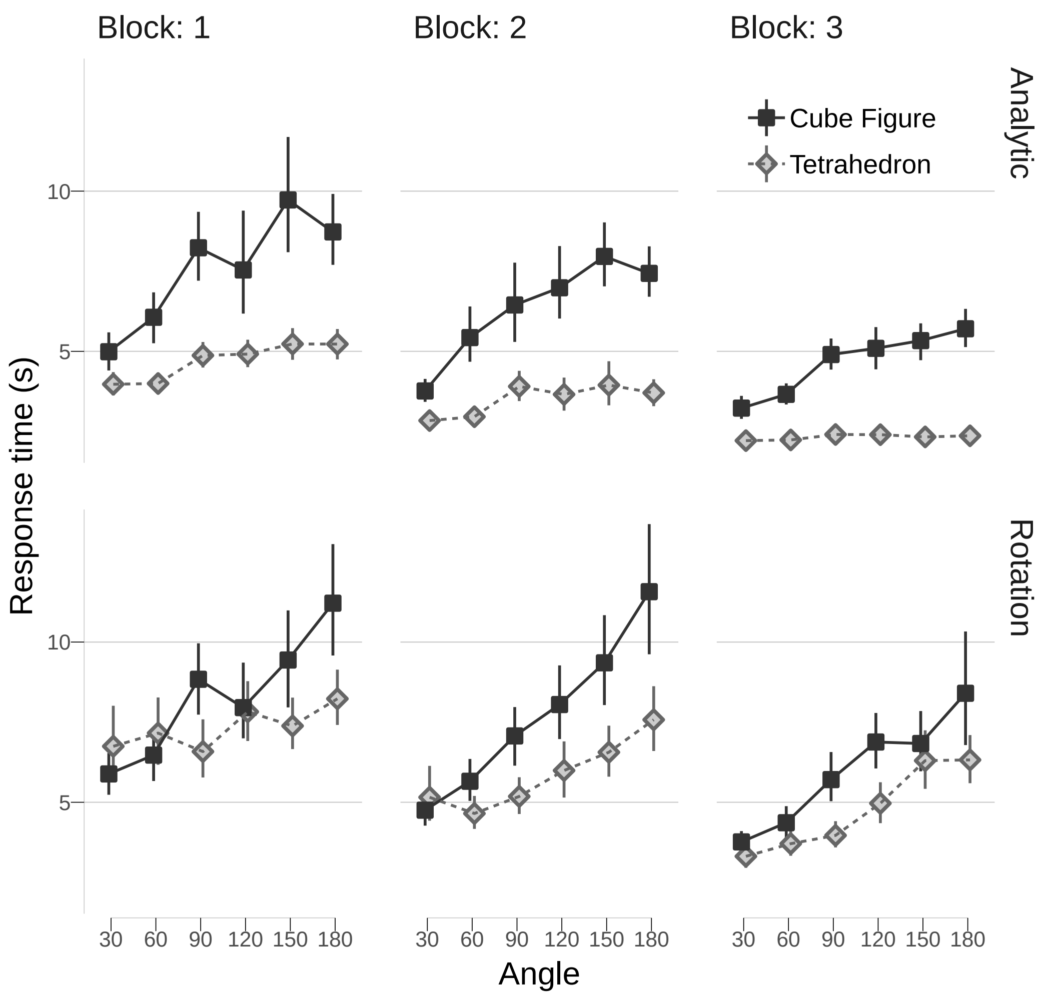

Figure 4.11: Response times as a function of angular disparity by stimulus type by trial block.

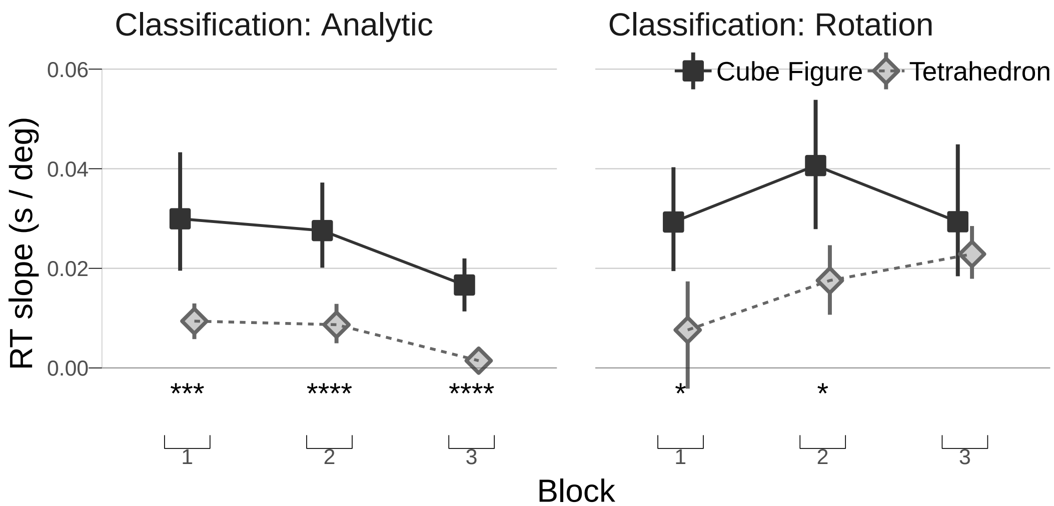

Figure 4.12: Response time slopes as a function of trial block by stimulus type by cluster class.

of trials. Participants classified as rotators had tetrahedron slopes that approached the slope for the cube figure stimuli. Importantly, tetrahedron slopes got smaller across trial blocks for those using the analytic strategy and got larger for those using the rotation strategy. Overall response time slopes are plotted in Figure 4.12.

Strategy use on error rate slope There were significant main effects of Stimulus type ($F[1, 62] = 45.30$, $MSE = 0.04$, $p < .001$, $\hat{\eta}^2_G = .169$) and Block ($F[1, 62] = 0.00$, $MSE = 0.05$, $p = .960$, $\hat{\eta}^2_G = .000$). These main effects were qualified by the two-way interaction between Block and Stimulus ($F[1.88, 116.75] = 8.64$, $MSE = 0.02$, $p < .001$, $\hat{\eta}^2_G = .028$). In general, participants classified as using an alternate strategy had tetrahedron slopes that approached zero during the third block of trials. Participants classified as rotators had tetrahedron slopes that approached the slope for the cube figure stimuli. In general, tetrahedron trial error rates were not affected by angular disparity compared to cube

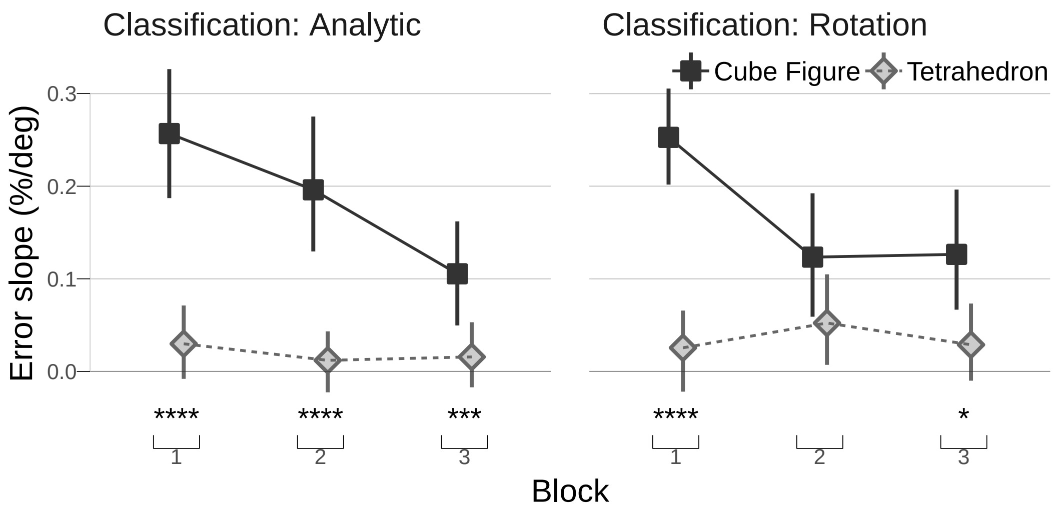

Figure 4.13: Error rate slopes as a function of trial block by stimulus type.

figure trials. See Figure **??**fig:02plot-errors) for overall error rate slopes by trial block.

Strategy use on response time slope distribution visualization Visual inspection of the change in slope distributions (degrees per second) in Figure **??**fig:02plot-dens) reveals participants became faster at the tetrahedrons only when using the analytic strategy.

In this experiment I determined that providing stereo viewing did not have an appreciable impact on strategy use nor performance during the task, and validated the stereo stimuli display apparatus. Self-reports of strategy as well as statistical clustering on behavioral data indicated that a subset of participants utilized a mental rotation strategy for both stimuli types while a larger group of participants utilized an alternate non-rotation cognitive strategy for the tetrahedral stimuli.

Next, I wanted to understand why certain individuals adopted the alternate strategy

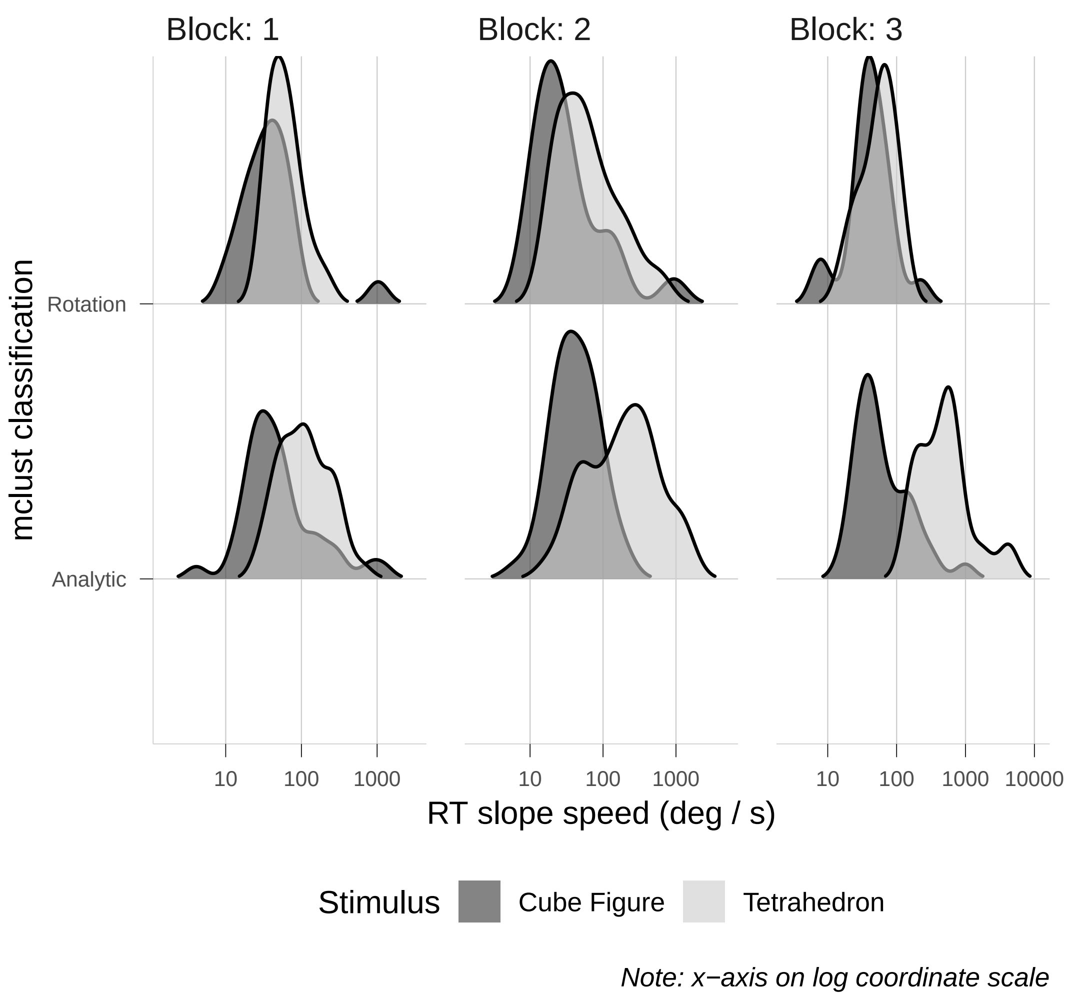

Figure 4.14: Density plots displaying the distribution of rotation rates (log scale) as a function of trial block by strategy classification.

while other did not. Can I identify then predict users who are likely to employ the alternate strategy based on pen-and-paper psychometric spatial ability measures?

Chapter 5

Experiment 3: Strategy and spatial ability

Introduction

Experiment 3 was conducted to examine the impact of spatial ability on performance and strategy use. Given the availability of an alternate strategy in the case of tetrahedron stimuli, we wanted to examine whether psychometric measures of ability differentially predicted strategy adoption. Mental rotation is one strategy for solving the Cube Comparison test, so we predicted that those with high scores should be more likely to employ a mental rotation strategy regardless of stimuli type. On the other hand, it could be that people with poor mental rotation ability search for and exploit alternate strategies in order to obviate use of mental rotation.

Gaining a better understanding of this connection should inform the relationship between strategy use and ability. Specifically, this experiment should shed light on whether spatial ability obviates the benefit of employing alternate strategies. Raw mental rotation ability should predict performance with both stimuli types, as mental rotation is available for both stimuli types. If ability predicts strategy adoption, then high ability individuals should lack an angular disparity effect for the tetrahedron stimuli. If no angular disparity effect is observed for low ability individuals, this would indicate that ability and strategy use are separable. If low ability individuals use the alternate strategy effectively, limitations of spatial ability could be circumvented by trained strategy use. In the context of STEM education and training, it is important to know if ability and strategy are separable. Elucidating potential differences in strategy among individuals with high and low spatial ability is important for instruction, especially if strategies employed by high achieving students can be learned by students with lower ability level.

Method

Design.

Experiment 3 had a similar experimental design as the second experiment without the between-subjects display format variable and the addition of three measures of ability as observed variables. The overall experimental design was a two (stimulus: cube figure or tetrahedron) by three (Block: first, second, third) by six (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) mixed-design. The dependent measures were response time (on correct *same* trials) error rate, as well as slope and intercept. We also employed full Bayesian data analysis using strategy classification as an observed quasi-dependent response variable predicted by the three ability measures.

Participants.

93 college students from the psychology subject pool at a research university participated in the study in return for course credit. None of the participants had previously studied organic chemistry. All participants had normal or corrected-to-normal vision. Three participants were excluded due to near chance performance with less than 60% accuracy rates overall. Of the 90 participants that met the criteria for inclusion, 48 were female. The mean age of participants was 19.3 years (SD = 1.1, range = 18, 22).

Materials.

The materials the same as in the previous experiments with the addition of three measures of ability: Cube Comparison, Paper Folding, and Abstract Reasoning factor-referenced cognitive tests from Ekstrom et al. (1976). Three measures of spatial ability were chosen to provide spatial skills assessments that could broadly be defined in terms of Harris, Newcombe, and Hirsh-Pasek's (2013) taxonomy of dynamic spatial transformations. The tests we chose covered Intrinsic-Dynamic (Paper Folding), Extrinsic-Static (Cube Comparison) as well as a more Extrinsic-Dynamic (Abstract Reasoning) spatial reasoning processes.

Cube Comparison Test

The Cube Comparison test consisted on two blocks of problems with 21 questions each. Each block had a time limit of three minutes. An example item from the test is shown in Figure 5.1 and instructions for the Cube Comparison Test were as follows:

> In this test, you will see two drawings of a cube. Assuming that no cube can

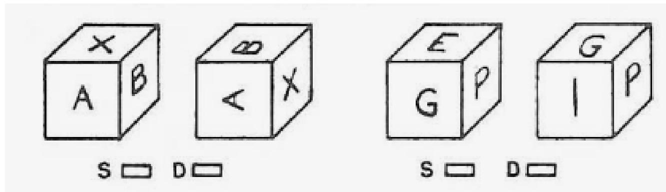

Figure 5.1: Example of two Cube Comparison test problems.

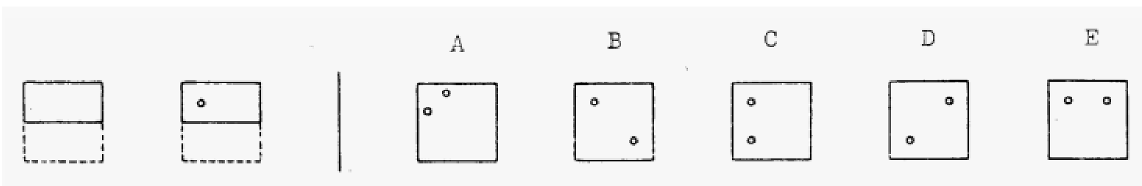

Figure 5.2: Example of a Paper Folding test problem.

> have two faces alike, you should indicate whether the two cubes can be of the same cube or if they are different cubes. Be sure to consider the orientation of the letters on a given face. No numbers, letters, or symbols appear on more than one face for a given cube.

Paper Folding Test

The Paper Folding Test consisted on two blocks of problems with ten questions each. Each block had a time limit of three minutes. Instructions for the Paper Folding Test were as follows:

> In this test, you will see drawings of a square sheet of paper that has two to three folds. In the final drawing of the folded paper, a hole is punched through all of the thickness of the paper at that point. Your job is to select one of the 5 drawings that shows how to paper would look if it was fully reopened.

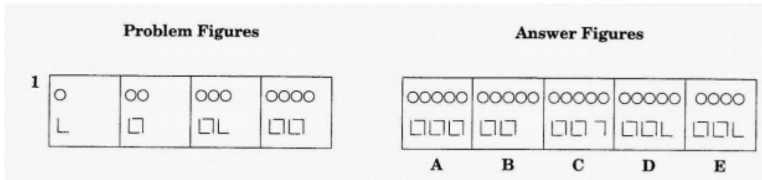

Figure 5.3: Example of an Abstract Reasoning test problem.

Abstract Reasoning Test

For the Abstract Reasoning test, participants were administered 40 problems with a time limit of 10 minutes. An example item from the test is shown in Figure 5.3 instructions for the Abstract Reasoning Test were as follows:

> In this test you will be solving problems with figures or designs. Each problem has four Problem Figures and five Answer Figures. The four Problem Figures form a series with something happening in each Problem Figure. Your task is to choose the answer figure that would be the next figure (fifth figure) in the series. After choosing, mark the space on your answer document for the answer you have chosen.

Procedure.

Groups of up to 3 individuals participated in the same windowless laboratory room as in previous experiments. After providing informed consent, participants verified with the experimenter that the image displayed was fusing properly for them. In contrast to Experiment 2, all participants viewed the experiment in stereo 3D. Any adjustments were completed by the experimenter at this time. Next, instructions were read aloud by the experimenter while participants followed along with written instructions displayed on the monitor. Next, participants completed four practice trials with feedback, and were instructed to work as quickly and accurately as possible through the self-paced experimental trials. After the first block, participants completed four practice trials with the other stimulus type. There were no practice trials for all subsequent blocks. After completing the experimental trials, participants completed the self-report strategy questionnaire before exiting to the main room in the lab to take the psychometric battery. The spatial battery took about 45 minutes to complete, afterwards participants were debriefed, thanked, and given credit for their participation.

Results and Discussion

Data screening

Screening for fast and careless respondents showed that there were no participants that had less than .5s response times on more than 20% of total trials. Responses faster than .5s (0.42% of total trials) were not included in the analysis. There were no outlying long responses.

Replication summary

Overall response time and error rate patterns are displayed in Figure 5.4. A 2 (Stimulus type: cube figure, tetrahedron) x 3 (trial Block: first, second, third) x 6 (angular disparity: 30°, 60°, 90°, 120°, 150°, 180°) within-design ANOVA was conducted on raw response times for correct non-mirror trials. Angular disparity ($F[3.52, 249.87] = 201.56$, $MSE = 7.06$, $p < .001$, $\hat{\eta}^2_G = .222$), Stimulus ($F[1, 71] = 95.97$, $MSE = 25.65$, $p < .001$, $\hat{\eta}^2_G = .123$), and Block ($F[1.77, 125.99] = 177.64$, $MSE = 12.34$, $p < .001$, $\hat{\eta}^2_G = .181$)

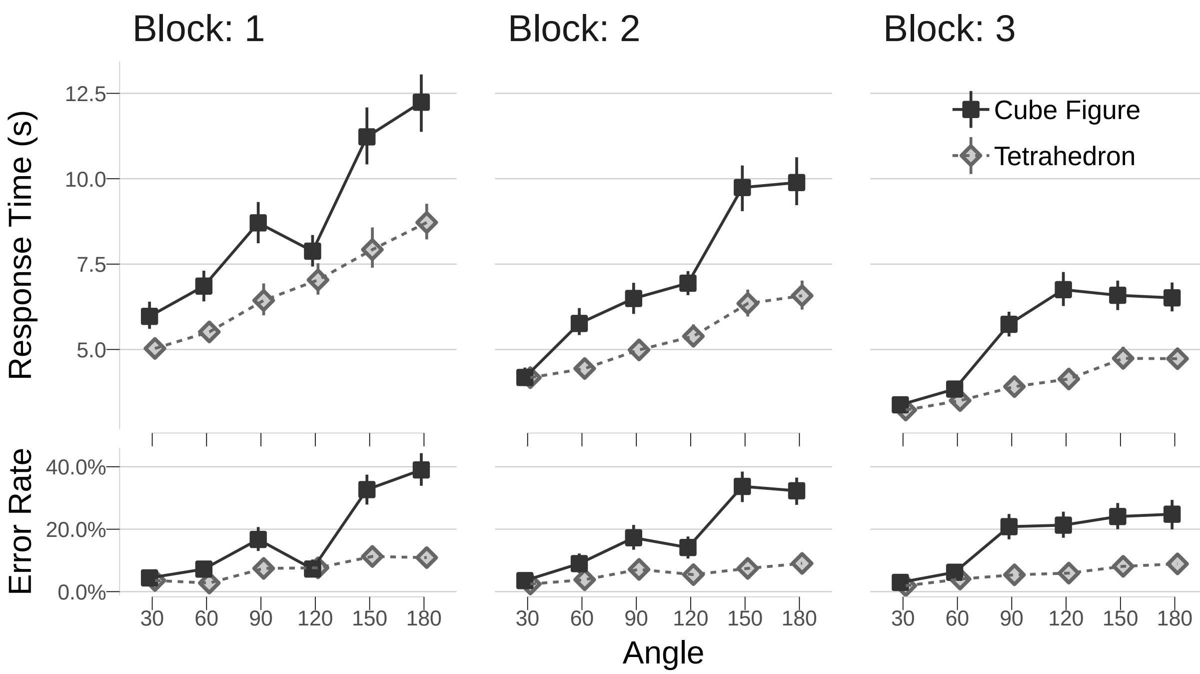

Figure 5.4: Response times and error rates as a function of angular disparity by trial block with 95% confidence intervals.

affected response time. A significant two-way interaction between Stimulus and Block ($F[3.8, 269.97] = 24.41$, $MSE = 6.40$, $p < .001$, $\hat{\eta}^2_G = .033$) was observed. These results replicated the trends observed in the first two experiments.

Table 5.1: Analysis of variance table for response time slopes on 'same' trials.

Effect	F	df1	df2	MSE	p	ges
Stimulus	53.06	1	89	0.00	$< .001$	.125
Block	34.52	1.57	139.33	0.00	$< .001$	.082
Stimulus × Block	2.44	1.54	137.28	0.00	.105	.008

Response time slope Response time slope was the dependent variable in a 2 (Stimulus type: cube figure, tetrahedron) x 3 (trial Block: first, second, third) ANOVA. Both main effects of Stimulus type ($F[1, 87] = 22.52$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .054$) and trial

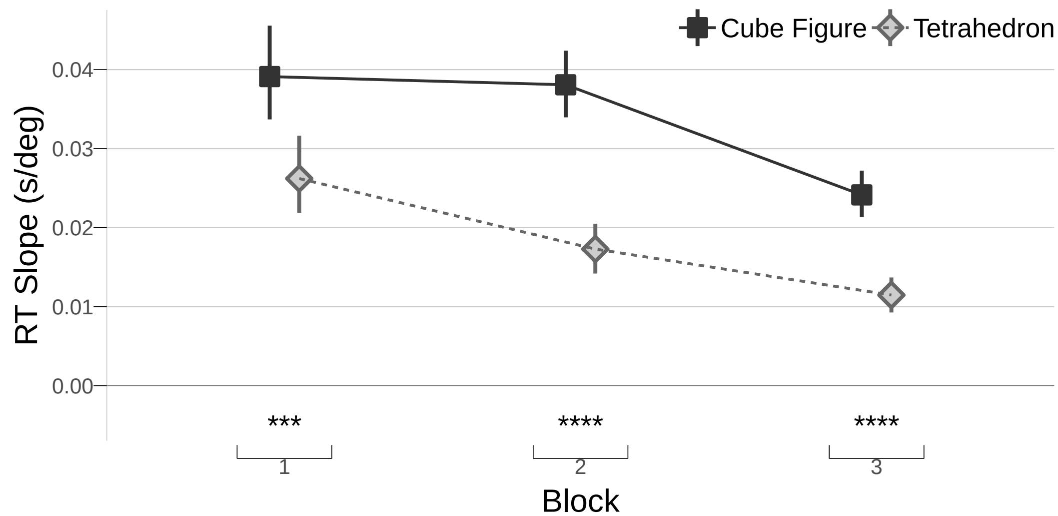

Figure 5.5: Response time slopes as a function of trial block with 95% confidence intervals.

Block ($F[1.9, 165.52] = 0.65$, $MSE = 0.02$, $p = .515$, $\hat{\eta}^2_G = .002$) reached significance. The Stimulus and Block interaction did not reach significance ($F[1.94, 168.54] = 1.91$, $MSE = 0.01$, $p = .152$, $\hat{\eta}^2_G = .005$). Overall, the response time slope analysis replicated the general trend observed in Experiment 2.

Table 5.2: Analysis of variance table for error rate slopes on 'same' trials.

Effect	F	df1	df2	MSE	p	ges
Stimulus	106.96	1	89	0.03	$< .001$	.214
Block	5.27	1.91	169.67	0.02	.007	.015
Stimulus $\times$ Block	3.27	1.93	171.91	0.01	.042	.008

Strategy use

Like in Experiment 2, strategy use class was assigned by using the same statistical clustering method on the distribution of individual response time slopes for the third block of tetrahedron trials.

Self-reported strategy performance

Two trained independent raters coded the strategy questionnaires to assign people as either rotators, alternate strategy users, or ambiguous. Inter-rater reliability was high (Cohen's Kappa $> .80$) and a third rater reconciled few disagreements, and three ambiguous cases were not included. Table 5.3 shows the overall mean performance of each strategy group based on the self-reported strategy questionnaire. In general, those who reported using the analytic strategy were faster and more accurate with the tetrahedron trials than those who did not report using the alternate strategy (rotators). Participants who reported using an analytic strategy had faster response times and fewer errors for tetrahedron trials on average. Overall, participants who reported using a rotation strategy completed the tetrahedron trials at a similar rate as the cube figure trials.

Table 5.3: Performance by self-reported strategy classification

Self-report	Stimulus	RT (s)	slope (deg/s)	Error rate	n
Analytic	Cube Figure	9.1	39	0.2	25
Analytic	Tetrahedron	7.3	127	0.08	20
Rotation	Cube Figure	8.8	52	0.1	18
Rotation	Tetrahedron	7.1	32	0.09	27

Model based clustering

As in Experiment 2, the goal of this analysis was to assign a predicted strategy classification to each participant. Response time slopes for the last block of tetrahedron trials for each participant were analyzed using the same method with Mclust (Scrucca et al., 2016). If multiple strategies are being used and one strategy is orientation independent, that group should have a distribution of slopes that are both smaller and less variant. A cluster of more positive slopes (angular disparity effect) could be interpreted as a group of participants using a mental rotation strategy. In other words, response time slopes should in general be shallower and less variant for participants that use an alternate non-rotation strategy compared to those who employ a mental rotation strategy. Figure **??** shows the estimated density plot a large left-skewed mode of slopes approaching zero as well as another more positive, smaller mode. Just as in the previous experiment, equal variance and unequal variance models were fit to the third block of tetrahedron trials. The Bayesian Information Criterion, based on the likelihood function and is penalized for model complexity, indicated that the two-component equal variance model produced the best fit.

The summary of the fitted uni-variate, equal variance Gaussian finite mixture model by the Expectation Maximization algorithm produced a clustering table with 63 participants in cluster 1 and 30 in cluster 2. Cluster 1 had response time slopes clustered around zero (M = 186.12 deg/s, S = 3.06e-05 s/deg) and cluster 2 had more positive slopes (M = 40.647 deg/s). The parametric bootstrap likelihood ratio test was used to determine the optimal number of mixture components for the uni-variate equal-variance model. The results showed that the single-component model could be rejected in favor of the two-component model with a likelihood ratio test score (LRTS) p-value of .001.

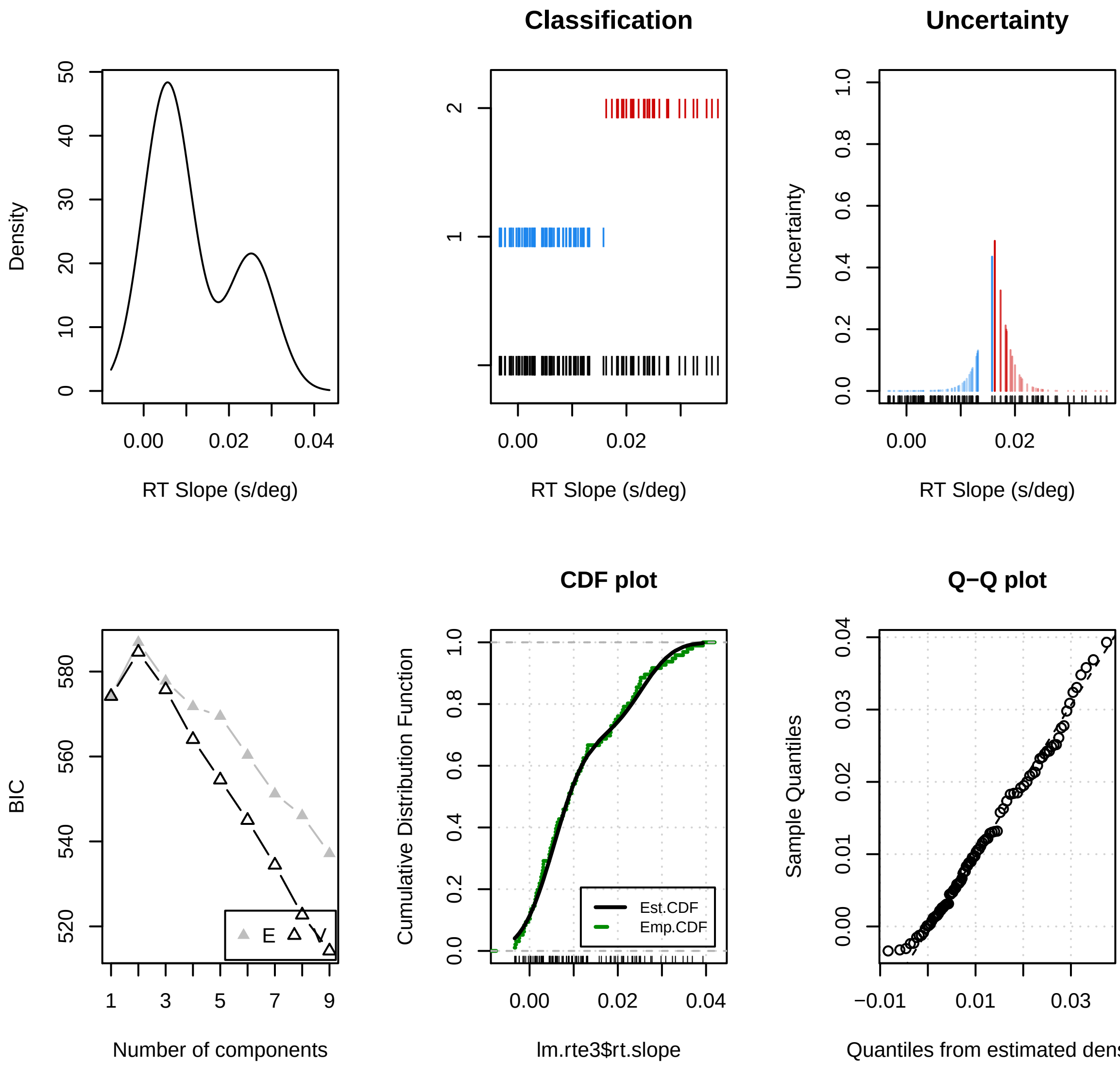

Figure 5.6: Estimated density plot (top left) and model-based clustering analysis of response time slopes. The preferred variable-variance two-component Gaussian finite mixture model had the highest Bayesian Information Criterion (bottom left).

Table 5.4: Clustering results table by self-reported strategy.

Self.report	cluster_1	cluster_2
Alternate	29	16
Rotation	9	36

The two-component model however could not be rejected in favor of the 3 component model ($LRTS = 7.82$, $p >.05$).

In contrast to Experiment 2, the difference in the Bayesian Information Criterion was negligible, indicating the equal variance model fit the data about as well as the unequal variance model.

Supervised classification

To evaluate the preferred model, we determined its predictive accuracy on labeled training data. Discriminant analysis was conducted to obtain the predictive error of the classifier. Results of the discriminant analysis are visualized in Figure 5.7. The model had a predictive error of 26% on the training data set. This could be interpreted as the classifier agreeing with the self-reported strategy 74% of the time. This seemed reasonable given the know unreliability of self-reports of cognitive strategy, especially if participants switch or use multiple strategies. For example, participants may not report their initial strategy they used or vice-verse. Self-reported versus cluster analysis derived strategy classifications are shown in Table 5.4. The model had a predictive error of 28% on the training data set. This again seemed reasonable, as it replicated the predictive error from the model in Experiment 2.

Cross validation In order to validate the predictive accuracy of the model, k-fold cross validation was conducted. 10-folds were randomly shuffled, and the predictive error

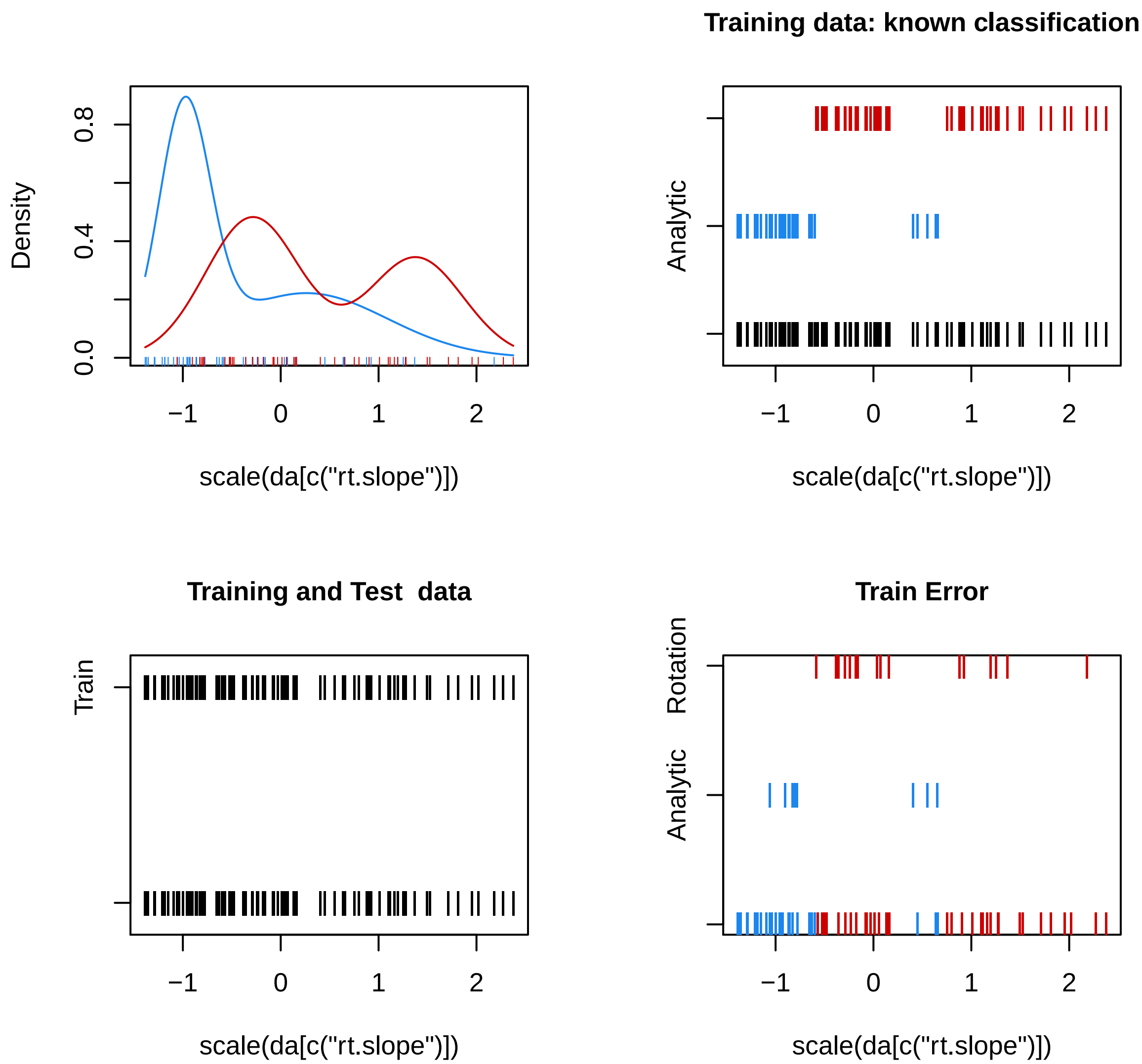

Figure 5.7: Discriminant analysis results summary. Top-left panel shows the two estimated distributions of response time slopes based on self-report. Known classifications (self-reported strategy) are displayed top-right. Bottom-right panel displays where the training errors occurred. The mixture-based discriminant analysis had a predictive error of 28%.

produced was *Error* = 33.3%, *SE* =4.48%. The k-fold cross validation results indicated the model was not over-fit, evidenced by similar predictive and training errors.

Performance by strategy

Response time slope by classification

Response time slopes for the table of assigned versus self-reported strategy class are shown in Figure 5.8. Participants who picked up on the analytic strategy had shallower slopes for the tetrahedron figures, and these slopes became flatter across blocks of trials. The participants classified as rotators had similar angular disparity effects for both the tetrahedron and cube figure stimuli.

Rotation Rates by Strategy Classification Another way to interpret this finding is to use the inverse of the response time slope to obtain estimated rotation rates in degrees per second. Plotting the distributions of rotation rates for each strategy class shows a clear separation between the cube figure and tetrahedron stimuli with increased practice. For example, in the third block of trials, some individuals were completing the tetrahedron stimuli trials at the equivalent rotation rates of 100-1000 deg/s and faster – impossibly faster than previously reported mental rotation rates. Both main effects of Stimulus type ($F[1, 88] = 34.88$, $MSE = 0.00$, $p < .001$, $\hat{\eta}^2_G = .084$) and strategy Classification ($F[1, 88] = 13.15$, $MSE = 0.00$, $p < .001$, $\hat{\eta}^2_G = .034$) reached significance. The two-way interaction between Block and strategy Classification did not reach significance ($F[1.55, 136.28] = 1.26$, $MSE = 0.00$, $p = .280$, $\hat{\eta}^2_G = .003$). In general, participants that were classified as using an alternate strategy had tetrahedron slopes that approached zero during the third block of trials. On the other hand, participants

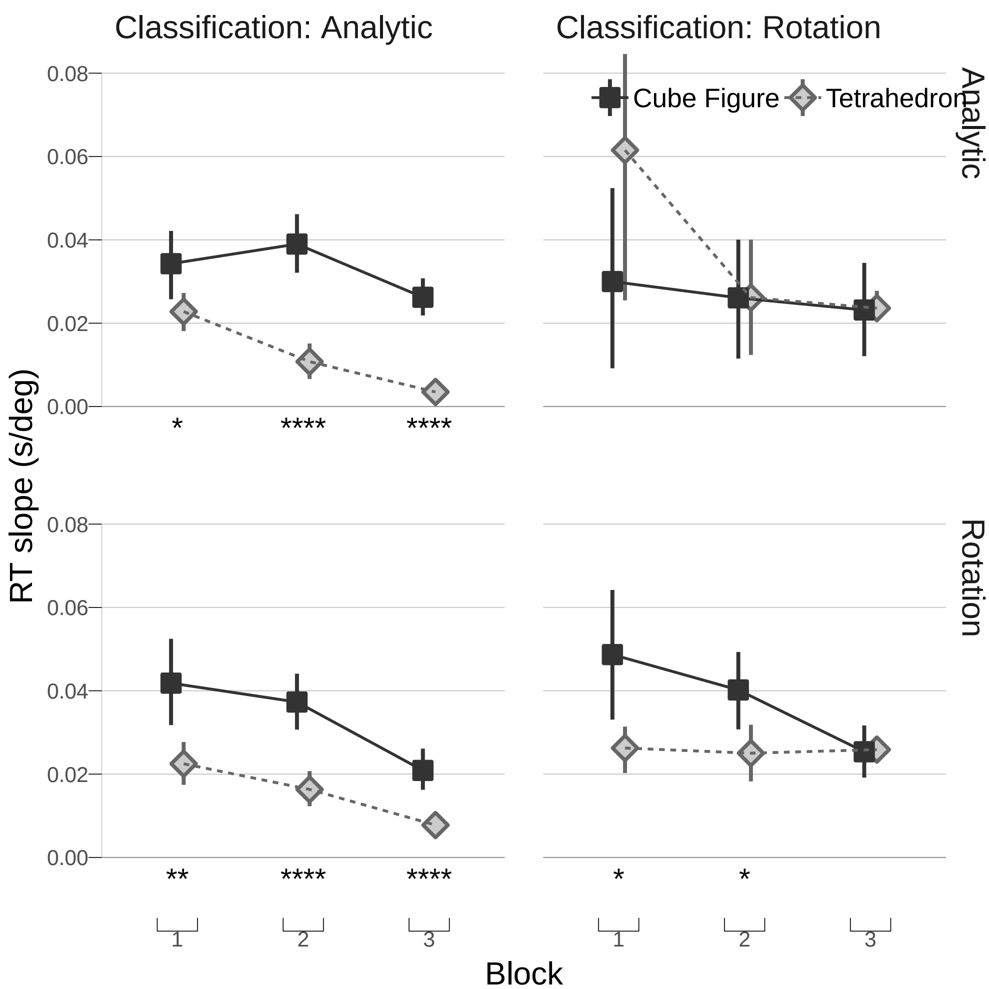

Figure 5.8: Response time slopes as a function of trial block by self-reported strategy by cluster-derived classification.

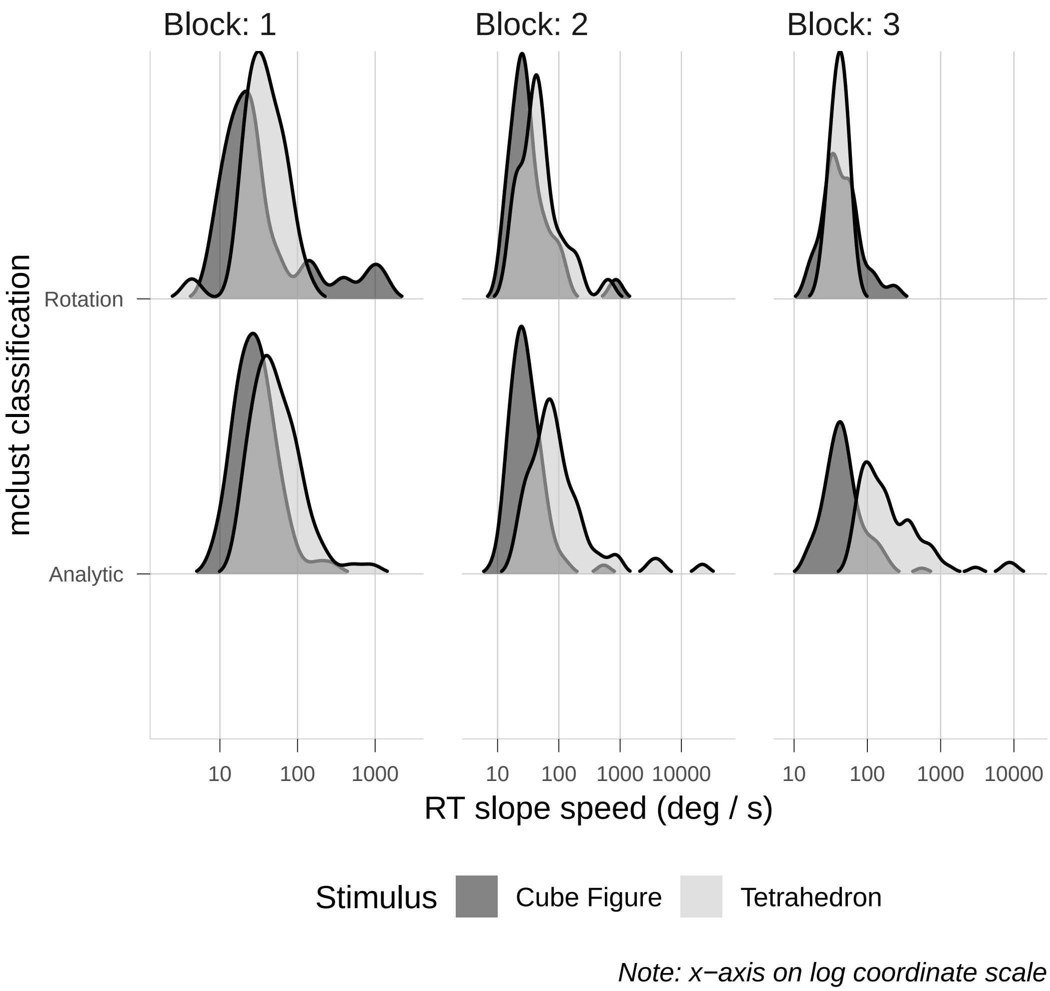

Figure 5.9: Density plots displaying the distribution of rotation rates (log scale) as a function of trial block by strategy classification.

classified as rotators had tetrahedron slopes that approached the slope for the cube figure stimuli.

Did ability affect performance and/or strategy use?

A central question in this study was to examine how ability related to performance as well as strategy use. As shown in Figure 5.10, there were significant moderate (all $r = .43$ - .54) correlations between each of the three ability measures. Importantly, there was a significant correlation between Cube Comparison and response time slope. A significant positive correlation between Paper Folding and error rate slope as well as a negative correlation between Abstract Reasoning and response time intercept were also observed.

Ability by self-reported strategy

Visual inspection of the distributions of spatial ability scores for the different strategy classifications in Figure 5.11 shows that in general people who reported using a rotation strategy scored higher on the Cube Comparison test. However, strategy was not associated with Abstract Reasoning or Paper Folding test scores.

Did ability predict strategy class?

To analyze test whether ability predicted strategy use, a Bayesian generalized linear model was fit to scaled ability measure scores to predict participant's strategy classification. Examination of the posterior distribution allows for effects to be interpreted in terms of likelihood and quantified uncertainty. The full Bayesian model was fitted using Markov Chain Monte Carlo (MCMC) with the R package rstanarm (Goodrich,

Figure 5.10: Correlation matrix of psychometric ability measures and task performance. Cells containing a 0 did not reach significance.

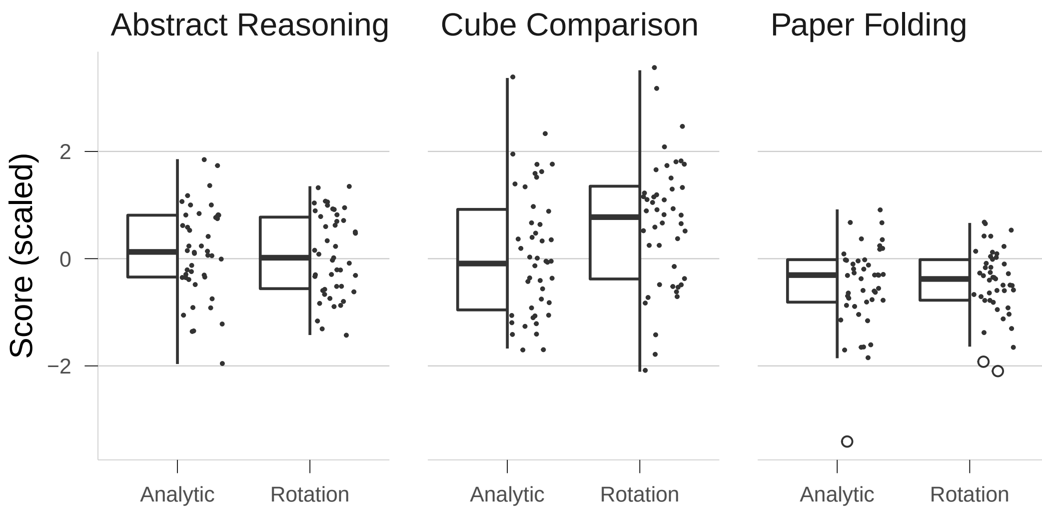

Figure 5.11: Distribution of scaled ability measure scores.

Gabry, Ali, & Brilleman, 2018). A Markov Chain Monte Carlo binomial (link = logit) model (4 chains, each with 1000 warm-up and 1000 post-warm-up iterations) predicted participants' classifier-assigned strategy. The outcome variable was parameterized such that positive coefficients relate to more positive slopes indicating the rotation strategy classification and negative coefficients relate to the analytic strategy classification.

The model's formula was specified:

```
Class ~ z.cct + z.pft + z.abs
```

The model's priors were set to be weakly informative:

```
~ normal (location = (0, 0, 0),
          scale = (2.56, 2.57, 2.54))
```

The results of the model were interpreted in the Bayesian framework by analyzing char-

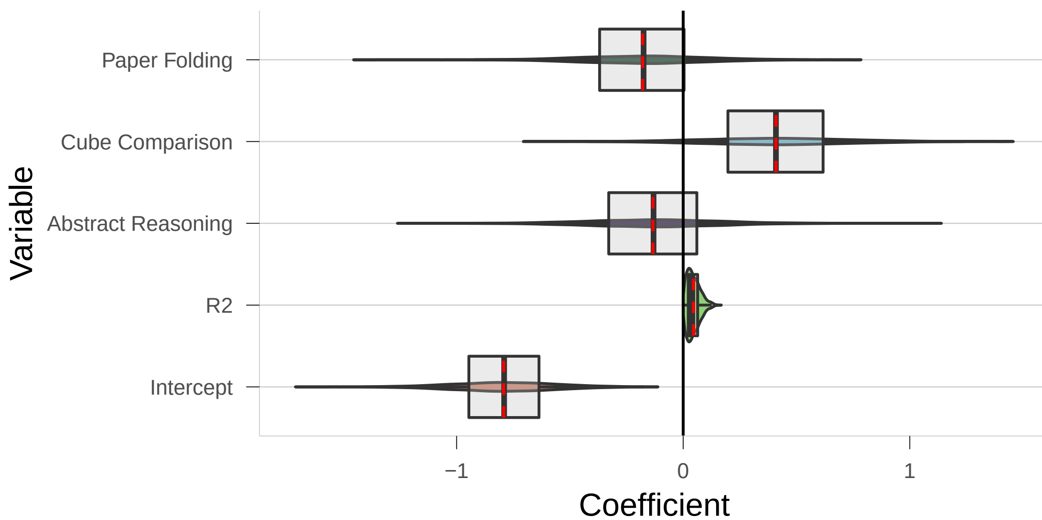

Figure 5.12: Bayesian generalized linear model results showing posterior distributions with 90% credible intervals. Positive coefficients represent the mental rotation strategy classification. People who scored higher on Cube comparison tended to employ the rotation strategy.

acteristics of the posterior distribution. Here we used the median for centrality and median absolute deviation (MAD) for dispersion. 90% *credible* intervals were calculated as estimates of where the true effect can actually exist with a probability of 90%. The maximum probability of effect (MPE) was computed. MPE is the probability that the median is in the same direction of the effect. In this analysis we considered effects with an MPE greater than 90% to be significant in the frequentist p-value sense.

Analyzing the results of the model, it had an explanatory power (R2) of about 4.00% (MAD = 0.029, 90% CI [0.0024, 0.088]). The intercept was -0.79 (MAD = 0.23, 90% CI [-1.17, -0.42]). Figure 5.12 plots the posterior distribution of each estimated predictor variable. The intercept was negative, which was expected as people were more likely to be classified as non-rotators overall. The main effect of Cube Comparison had a probability

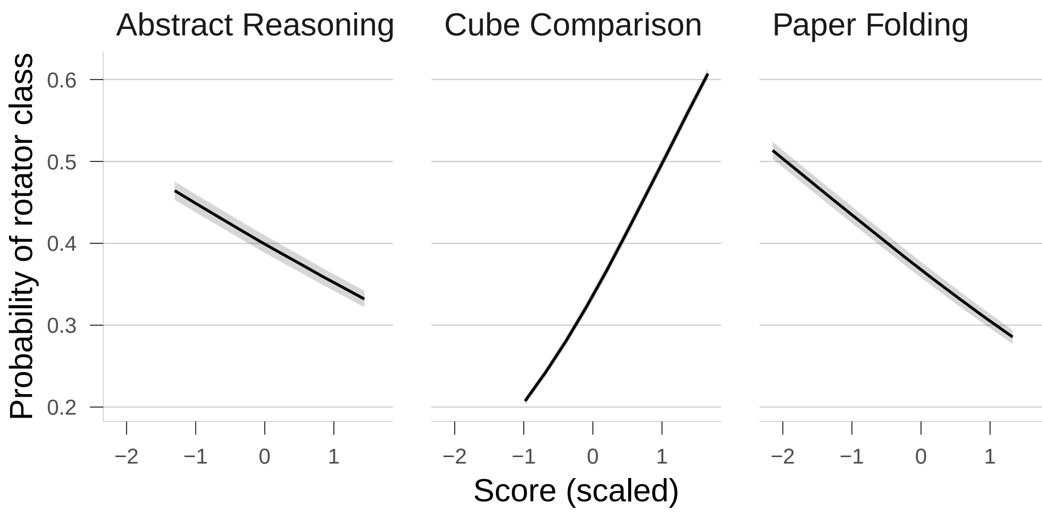

Figure 5.13: Logistic regression results showing the main effects of scaled psychometric scores on strategy classification. Participants who scored higher on Cube Comparison tended to be classified as using the rotation strategy.

of 90.70% of being positive (Median = 0.41, MAD = 0.31, 90% CI [-0.097, 0.91], Overlap = 51.94%). In other words, higher Cube Comparison scores was associated with rotator strategy classification. The effect of Paper Folding had a probability of 74.58% of being negative (Median = -0.17, MAD = 0.28, 90% CI [-0.63, 0.27], Overlap = 74.86%), and its effect is considered marginal without more data. The effect of Abstract Reasoning had a probability of 68.40% of being negative (Median = -0.13, MAD = 0.29, 90% CI [-0.63, 0.30], Overlap = 81.82%), and was also considered marginal. As shown in Figure 5.13, participants with higher Cube Comparison scores were more likely to be classified as rotators. Mental rotation is an effective strategy for solving the Cube Comparison test, while Paper Folding and Abstract Reasoning are not solvable using mental rotation only. Paper Folding and Abstract Reasoning had smaller trends in the opposite direction with negative coefficients indicating greater likelihood to adopt alternate strategy use.

However, these effects were too uncertain (MPE < .8) to consider probable without more evidence.

Did ability predict response time slope?

To analyze test whether spatial ability affected response time slopes for each stimulus type, Bayesian generalized linear models were fit to scaled ability measure scores. A Gaussian (link = identity) model (4 chains, each with 1000 warm-up iterations and 1000 post-warm-up iterations) estimated the intercept, the main effect of stimulus type, 3 main effects of ability measure, and 3 interactions between ability measure and stimulus type. The model's formula was specified:

```
rt.slope ~ Stimulus * (z.cct + z.pft + z.abs))
```

The model's priors were set to be weakly informative:

```
~ normal (location = (0, 0, 0, 0, 0, 0, 0),
          scale = (0.036, 0.036, 0.037, 0.036, 0.052, 0.052, 0.051))
```

The model had explanatory power (R2) of about 24.32% (MAD = 0.045, 90% CI [0.17, 0.31], adj. R2 = 0.17). The intercept was at 0.024 (MAD = 0.0013, 90% CI [0.022, 0.026]). The effect of tetrahedron stimuli had a probability of 100% of being negative, i.e. more speeded non-rotation trials (Median = -0.013, MAD = 0.0018, 90% CI [-0.016, -0.0097], Overlap = 0.48%). As shown in Figure 5.14, High Abstract Reasoning ability predicted more positive slopes for cube figure stimuli and more flat slopes for tetrahedron stimuli. In other words, high Abstract Reasoning predicted more positive slopes (indication of mental rotation) for cube figures, and more negative slopes (indication of non-rotation strategy) for tetrahedron stimuli trials. The main effect of Abstract Reasoning had a

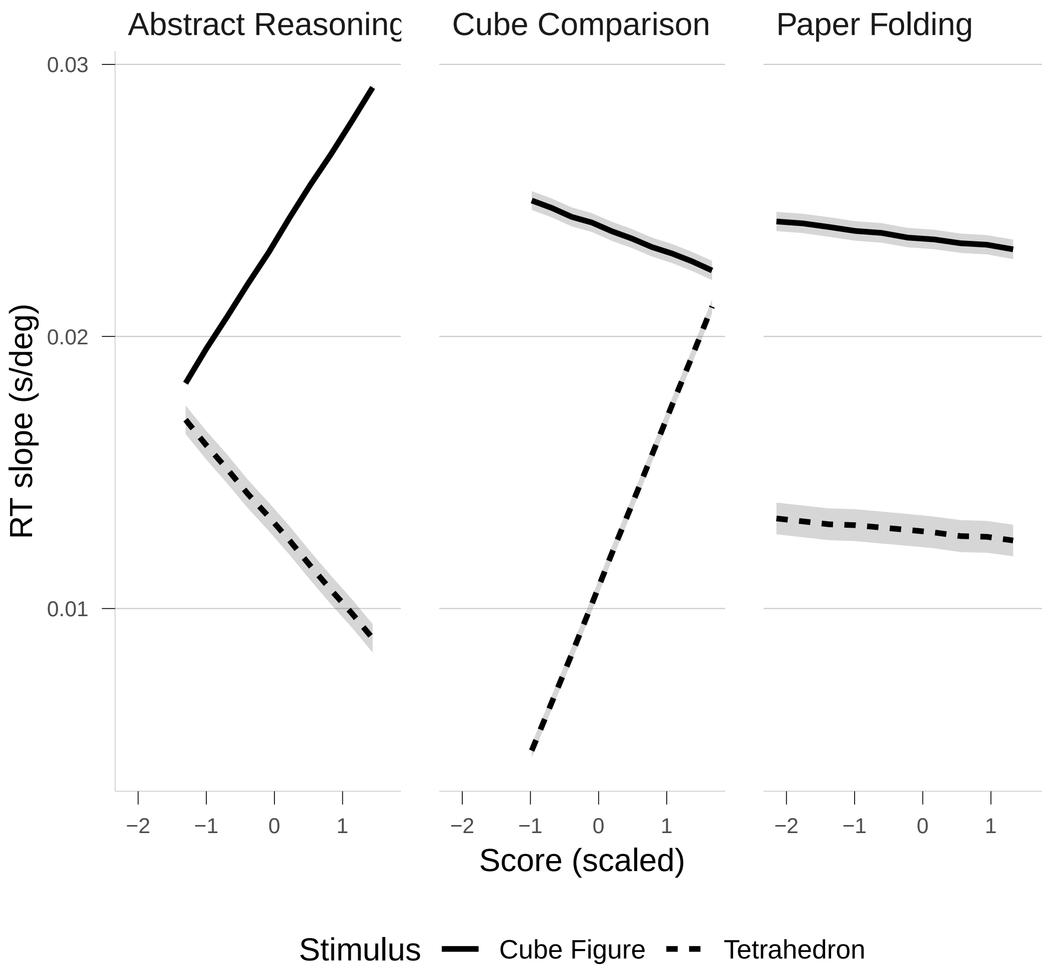

Figure 5.14: Regression results showing the effect of each ability measure on response time slope by stimulus type.

probability of 90.45% of being positive (Median = 0.0022, MAD = 0.0017, 90% CI [-0.00059, 0.0049], Overlap = 59.45%). The effect of Abstract Reasoning was qualified by an interaction with tetrahedron stimuli trials with probability of 95.50% of being negative (Median = -0.0039, MAD = 0.0023, 90% CI [-0.0077, 0], Overlap = 46.19%). Participants with low abstract reasoning had similar slopes for tetrahedrons and cube figures and high abstract reasoning individuals had large slopes for cube figures and much smaller slopes for tetrahedrons.

Cube Comparison also had a differential effect on response time slope, as seen in the second panel of Figure 5.14. Higher Cube Comparison scores predicted more positive slopes specifically on the cube figure stimuli. In other words, Cube Comparison predicted more positive slopes (indication of mental rotation) for the cube figures but not for the tetrahedron stimuli. The effect of Cube Comparison had a probability of 61.85% of being negative (Median = -0.00051, MAD = 0.0017, 90% CI [-0.0033, 0.0024], Overlap = 90.99%). The main effect of Cube Comparison was qualified by the interaction with stimulus type: The interaction between the effect of tetrahedron stimuli trials and Cube Comparison had a probability of 94.95% of being positive (Median = 0.0041, MAD = 0.0024, 90% CI [0, 0.0082], Overlap = 46.65%). People who scored low on cube comparison had small slopes for the tetrahedrons, while high cube comparison predicted large slopes for tetrahedrons.

As shown in the right panel of Figure 5.14, the effect of Paper Folding had a probability of 52.98% of being negative (Median = -0.00013, MAD = 0.0016, 90% CI [-0.0028, 0.0025], Overlap = 97.05%) and the interaction between stimulus type and Paper Folding had a probability of 50.22% of being positive (Median = 0, MAD = 0.0023, 90% CI [-0.0037, 0.0036], Overlap = 99.74%). Neither of these effects would be considered significant.

In general, participants with higher Cube Comparison scores were more likely to be classified as rotators. Rotators stuck with the mental rotation approach to solving both the classic cube figure stimuli trials as well as the molecule like tetrahedral trials. Abstract Reasoning ability was predictive of alternate non-rotation strategy use and led to improved performance compared to participants who used mental rotation for both stimuli types. Critically, we found that people with higher Cube Comparison scores were more likely to be assigned to the rotator strategy class.

Overall, this study showed how individual differences in ability differentially impacted propensity to utilize different strategies. Further, this study also highlights the importance of abstract reasoning ability for optimal strategy selection and utilization for spatial tasks with alternate solution paths.

Chapter 6

General Discussion

Over three experiments I explored and delineated the contributions of task, stimulus, and practice to mental rotation performance. Further, I was particularly interested in individual differences in cognitive strategy, heuristic use, and how ability moderated strategy use.

In Experiment 1, I compared performance on the mental rotation test with standard Shepard and Metzler (1971) stimuli versus tetrahedral models like those found in introductory organic chemistry. The time needed to decide whether the two objects are the same or mirror images of each other increased linearly with angular disparity on picture- and depth-plane trials. On aggregate, it appeared that depth-plane tetrahedron trials were being completed faster and more accurately than cube figure trials. However due to the identical structure of the tetrahedron stimuli (they only varied in the arrangement of colors), an alternate strategy to mental rotation was available for tetrahedron trials rotated about the depth-plane axis. People who picked up on this alternate strategy employed a heuristic of simply judging whether the inner (i.e. adjacent) colors across stimuli pairs was either the same color or different.

Visualizing individual differences in slope by stimuli type revealed a subgroup of participants who had slopes of similar values regardless of stimuli, and a larger group of people who had much flatter slope coefficients for the molecule like stimuli. This effect was only apparent on depth-plane rotation trials. Users who picked up on and employed the non-rotation heuristic were predicted to perform the task faster and more accurately than users who employed a mental rotation strategy. The next two experiments replicated the finding that a subset of participants picked up on the alternate non-rotation strategy. Gaussian mixture modeling on individual response time slopes revealed that the two-cluster model fit the data best in both Experiments Two and Three. In each experiment, about 70% of people were clustered into the class with slope coefficients approaching zero (alternate strategy users), and the remaining were clustered in the class with larger slopes that were indicative of the mental rotation process.

Administering Shepard and Metzler's (1971) mental rotation test using stimuli common in STEM education may have introduced ambiguous demand characteristics for participants familiar with the experimental design and these participants may have assumed the experimenter expected participants to use mental rotation. Students from the psychology subject pool were likely familiar with the methods and results of the original mental rotation experiments from introductory psychology instruction. This may have caused some people to assume that the experimenter expected them to perform mental rotation across all trials, despite there not being any mention of strategy in the instructions. The introduction of an alternate strategy for the tetrahedron trials caused people to bifurcate in their strategy use. When the alternate strategy was available on the tetrahedron trials, the majority exploited the alternate strategy while about 30% continued using mental rotation.

Why did one third of participants continue using mental rotation despite the availability of a more efficient and less error prone alternate strategy? Participants that utilized mental rotation for the tetrahedrons took longer to respond and made more errors than most participants who switched to the non-rotation strategy on the tetrahedron trials. Angular disparity did not affect performance of those who used the alternate strategy, so how should performance be evaluated in cases such as these? The optimal solution would be to employ the faster and more accurate non-rotation strategy as soon as it becomes apparent in the context of the problem at hand. However, without explicit instructions or cueing the of the alternate strategy, it could be expected that these participants simply stuck with mental rotation because the strategy transfers across stimulus types.

In Experiment 3, I examined how individual differences in ability moderated cognitive strategy utilization. Participants who used mental rotation tended to score higher on Cube Comparison than people who used the non-rotation strategy. The two strategy groups did not differ on the Paper Folding or Abstract Reasoning tests. Higher scores on the Cube Comparison Test predicted larger angular disparity effects on tetrahedron trials, indicating use of mental rotation. Notably, participants with the greatest difference in response time slopes by stimuli type tended to score higher on the Abstract Reasoning test. This indicated that the fastest and most accurate participants were able to flexibly switch between the non-rotation strategy for tetrahedron stimuli and mental rotation for the cube figure stimuli.

Mental rotation is the first strategy used during the task for the tetrahedron trials. In a factor analysis of pen-and-paper spatial ability measures, Hegarty and Waller (2005) found that the Mental Rotation Test (Vandenberg & Kuse, 1978) and the Cube Comparison test had roughly equivalent moderate to large factor loadings on mental rotation ability. In general, people who were good at Cube Comparison continued using mental rotation on all trials regardless of stimulus type. If mental rotation comes naturally to an individual there is likely less motivation to explore alternative solution paths in the problem space. Alternately, these findings could also be interpreted to as users over applying their preferred process-based cognitive strategy at the expense of response time and accuracy.

Scoring higher on Abstract Reasoning differentially predicted larger slopes for the cube figure stimuli and flatter slopes for the tetrahedron stimuli. The present data suggest that the faster response times (flatter slopes) observed on tetrahedron trials of high abstract reasoning ability people did not reveal an angular disparity effect, but rather was indicative of a completely different mental operation that obviated spatial transformation. Researchers studying chronometric mental rotation with different stimuli should use caution when interpreting slope coefficients across stimuli types that may introduce alternate solution pathways that could afford abstraction of simpler rule-based strategies. Note that if I were to average over individual participant slopes, this would cause the misinterpretation that people are faster at mental rotation with the simpler tetrahedral stimuli, when in fact most people were not using a mental rotation strategy at all. For this reason, it is important to examine individual differences in slopes to ensure that differences in aggregate performance have not arisen from a bifurcation of cognitive strategies employed by individual participants. Overall, this work demonstrates how the introduction of strategic ambiguity in spatial tasks may have differential effects on cognitive strategy use, rather than mental rotation performance, depending on individual ability level.

One limitation of the study was that only simple tetrahedral figures were tested. Often organic chemistry representations include hydrogen atoms on sub-groups, so these stimuli were slightly less complex than a typical tetrahedral molecule found in introductory organic chemistry. Future work should investigate whether adding components of sub-groups affects cognitive strategy use. Further, the fact that the alternate strategy was only available for depth-plane rotations about the horizontal axis indicated that the crucial effect arose from a combination of factors that could be a reason why some people did not switch to using it. For picture plane rotations I found that people used mental rotation for both tetrahedron and cube figure stimuli.

In relation to issues surrounding how to best train and improve spatial thinking for STEM education, these findings highlight the important skill of abstracting away spatial processing problems to non-spatial analytic judgments. Opportunities to abstract away spatial processing may arise from planes of symmetry inherent to molecular geometry, or other spatial geometries more broadly. The present work suggests that abstract reasoning ability is important for meta-cognitive judgments regarding whether to employ mental rotation versus the alternate strategy. These findings fit nicely with Stieff's (2012) finding that beginning organic chemistry students relied on mental rotation while domain experts were able to abstract away extraneous spatial information and relied on heuristics that instead require domain knowledge. The present work also integrates well with the results of a classroom training study by Stieff, Dixon, Ryu, Kumi, and Hegarty (2014) that administered three different interventions: imagistic training, analytic training, or a combination of both imagistic and analytic strategy training. They found that administering the combined training intervention eliminated the sex difference on the post-test and class achievement. Analytic, non-rotation strategy training improved post-test performance as well as course achievement for female students. Importantly, the present research demonstrates that people who used the more efficient alternate strategy had better performance on novel STEM stimuli, and this was not dependent on mental rotation ability. Instead, the present research highlights questions of the efficacy of generalized spatial skills training versus alternate approaches such as teaching analytic strategies. In the broader context of spatial instruction for STEM domains, designers of training interventions should focus on training representational competence– the ability to relate multiple representations to with domain specific heuristics. This work helps build on evidence that training alternate strategies for spatial problems in STEM could enrich spatial training interventions.

References

Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1005–1019.

Carroll, J. B. (. B. (1993). *Human cognitive abilities : A survey of factor-analytic studies*. Cambridge ; New York: Cambridge University Press.

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20–43.

Cooper, L. A., & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 503.

Ekstrom, R. B., Dermen, D., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests* (Vol. 102). Educational testing service Princeton, NJ.

Gardony, A. L., Eddy, M. D., Brunyé, T. T., & Taylor, H. A. (2017). Cognitive strategies in the mental rotation task revealed by eeg spectral power. *Brain and Cognition*, *118*, 1–18.

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating" rotators" from" nonrotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research*, *41*(3), 261–293.

Gluck, J., & Fitting, S. (2003). Spatial strategy selection: Interesting incremental information. *International Journal of Testing*, *3*(3), 293–308.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). Rstanarm: Bayesian applied regression modeling via Stan. Retrieved from http://mc-stan.org/

Harris, J., Newcombe, N. S., & Hirsh-Pasek, K. (2013). A new twist on studying the development of dynamic spatial transformations: Mental paper folding in young children. *Mind, Brain, and Education*, *7*(1), 49–55.

Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review*, *25*(3), 1212–1219.

Hegarty, M., Stieff, M., & Dixon, B. L. (2013). Cognitive change in mental models with experience in the domain of organic chemistry. *Journal of Cognitive Psychology*, *25*(2), 220–228.

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. *The Cambridge Handbook of Visuospatial Thinking*, 121–169.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*(2), 137.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 1479–1498.

Maeda, Y., & Yoon, S. Y. (2013). A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualization tests: Visualization of rotations (psvt: R). *Educational Psychology Review*, *25*(1), 69–94.

Metzler, J., & Shepard, R. N. (1974). Transformational studies of the internal representation of three-dimensional objects.

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted stem undergraduates? *Learning and Individual Differences*, *26*, 141–152.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *The R Journal*, *8*(1), 289.

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.

Shepard, R. N. (2004). How a cognitive psychologist came to seek universal laws. *Psychonomic Bulletin & Review*, *11*(1), 1–23.

Shepard, R. N., & Cooper, L. A. (1986). *Mental images and their transformations.* The MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Small, M. Y., & Morton, M. E. (1983). Research in college science teaching: Spatial visualization training improves performance in organic chemistry. *Journal of College Science Teaching*, *13*(1), 41–43.

Sorby, S. A. (2009). Developing 3-d spatial visualization skills. *Engineering Design Graphics Journal*, *63*(2).

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, *26*, 20–29.

Stieff, M. (2011). When is a molecule three dimensional? A task-specific role for imagistic reasoning in advanced chemistry. *Science Education*, *95*(2), 310–336.

Stieff, M., Dixon, B. L., Ryu, M., Kumi, B. C., & Hegarty, M. (2014). Strategy training eliminates sex differences in spatial problem solving in a stem domain. *Journal of Educational Psychology*, *106*(2), 390–402.

Stieff, M., & Raje, S. (2010). Expert algorithmic and imagistic problem solving strategies in advanced chemistry. *Spatial Cognition & Computation*, *10*(1), 53–81.

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

Stieff, M., & Uttal, D. (2015). How much can spatial training improve stem achievement? *Educational Psychology Review*, *27*(4), 607–615.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and stem education-chapter four:When, why, and how?: When, why, and how? *Psychology of Learning and Motivation*, *57*, 147–181.

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.

Uttal, Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current*

Directions in Psychological Science, *22*(5), 367–373.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. doi:10.2466/pms.1978.47.2.599

Voyer, D., Voyer, S., & Bryden, M. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. doi:10.1037/0033-2909.117.2.250

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250.

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.

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420.

Wu, H.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492.