Strategy Selection versus Flexibility: Using Eye-trackers to Investigate Strategy Use during Mental Rotation

Alina Nazareth[1], Rebecca Killick[2], Anthony S. Dick[3] and Shannon M. Pruden[3]

[1]Temple University

[2] Lancaster University

[3]Florida International University

Corresponding Author:
Dr. Alina Nazareth
Temple University
1701 N. 13th Street
Philadelphia, PA 19122
E-mail: alina.nazareth@temple.edu

Abstract

Spatial researchers have been arguing over the optimum cognitive strategy for spatial problem-solving for several decades. The current paper aims to shift this debate from strategy dichotomies to strategy flexibility — a cognitive process, which although alluded to in spatial research, presents practical methodological challenges to empirical testing. In the current study, participants' eye movements were tracked during a mental rotation task (MRT) using the Tobii X60 eye-tracker. Results of a latent profile analysis, combining different eye movement parameters, indicated two distinct eye-patterns — fixating and switching patterns. The switching eye-pattern was associated with high mental rotation performance. There were no sex differences in eye-patterns. To investigate strategy flexibility, we used a novel application of the changepoint detection algorithm on eye movement data. Strategy flexibility significantly predicted mental rotation performance. Male participants demonstrated higher strategy flexibility than female participants. Our findings highlight the importance of strategy flexibility in spatial thinking and have implications for designing spatial training techniques. The novel approaches to analyzing eye movement data in the current paper can be extended to research beyond the spatial domain.

Keywords: cognitive, sex differences, spatial, eye-movement, changepoint analysis

Strategy Selection versus Flexibility: Using Eye-trackers to Investigate Strategy Use during Mental Rotation

Mental rotation is the ability to mentally transform 2D and 3D objects in space (Shepard & Metzler, 1971). It is a fundamental spatial thinking skill and is used in everyday tasks like fitting luggage in the trunk of a car and rearranging furniture. Yet, there are large individual as well as sex differences in favor of men noted in the mental rotation literature (e.g., Halpern, 1992; Nazareth, Hererra & Pruden, 2013; Voyer, Voyer & Bryden, 1995). Mental rotation tasks like the assembled cubes task in Shepard & Metzler (1971) have been extensively used to study mental rotation ability in laboratories (see Figure 1). Mental rotation ability, among other spatial thinking skills, has been linked to entry into and success in Science, Technology, Engineering and Mathematics (STEM) fields (Delgado & Prieto, 2004; Guillot, Champely, Batier, Thiriet, & Collet, 2006; Hoyek et al., 2009; Lubinski, 2010; Wai, Lubinski & Benbow, 2009). This appreciation for the extent to which spatial thinking is central to educational and occupational settings (Lubinski, 2010) has led to a renewed interest in the field. A recent meta-analysis demonstrating the ability to improve spatial performance through training (Uttal, et al., 2013) increases the importance of understanding the sources of individual differences with the ultimate aim to design effective training techniques. One plausible explanation for individual and/or sex differences in mental rotation performance could be the differences in the cognitive process of strategy selection —an individual's preferential problem-solving style (Butler et al., 2006; Heil & Jansen-Osmann, 2008; Stieff, Dixon, Ryu, Kumi, & Hegarty, 2014).

Strategy selection pertaining to spatial reasoning has been extensively studied in the mental rotation literature (Hegarty, 2010; Khooshabeh, Hegarty, & Shipley, 2013; Kozhevnikov, Hegarty, & Mayer, 2002; Meneghetti, Cardillo, Mammarella, Caviola, & Borella, 2016; Shepard & Metzler, 1971; Ter Horst, Jongsma, Janssen, van Lier, & Steenbergen, 2012). For several decades' researchers investigating strategy selection during spatial thinking tasks have attempted to operationalize the strategies that participants' use, but an inconsistent vocabulary has been employed. For example, Shepard and Metzler (1971; Shepard & Metzler, 1988) described a cognitive strategy during mental transformation akin to the physical rotation of objects. In subsequent decades, a variety of strategy dichotomies including holistic/piecemeal (Khooshabeh et al., 2013), egocentric/allocentric (Ganesh, van Schie, Cross, de Lange, & Wigboldus, 2015), mental imagery/analytic (Stieff et al., 2014), flipping/spinning (Kanamori & Yagi, 2002), verbalizer/visualizer (Kozhevnikov et al., 2002) have been defined. Specific to the mental rotation literature, holistic and piecemeal strategies are two of the most frequently investigated strategies (Heil & Jansen-Osmann, 2008; Kail, Carter, & Pellegrino, 1979; Khooshabeh et al., 2013). A holistic strategy involves mentally rotating an object as a whole and encoding the complete spatial information of the presented object. For example, in Figure 1, the object on the right can be holistically rotated along an imaginary vertical axis for comparison with the object on the left. A *piecemeal* strategy, on the other hand, involves the breaking down of the object into several pieces and encoding only partial spatial information of the presented object. In the *piecemeal* strategy, an individual may mentally rotate one piece in congruence with the comparison object and then apply the same rotation to the other parts of the object to see if they match (Khooshabeh et al., 2013). For example, in Figure 1, one piecemeal heuristic would be to break the object on the right into 4 independent arms (pieces) for rotation for comparison with the 4 corresponding arms (pieces) of the object on the left. More recently, Xu and Franconeri (2015) demonstrated a third possible strategy in which participants tracked the top (quadrants) of objects and thus encoded and rotated parts of objects and not the whole object as indicated by a holistic strategy.

The field of spatial research is divided on which of these two strategies — holistic or piecemeal —enhances mental rotation performance. After decades of research on mental rotation tasks varying in stimuli, design and complexity, there is research supporting both sides of the argument. Supporting the holistic strategy, Heil and Jansen-Osmann (2008) found that males used a holistic strategy during mental rotation which enabled them to maintain their response times (RT) even with increasing stimulus complexity. Similarly, Khooshabeh and colleagues (2013) found that good performers showed a preference for a holistic strategy and were therefore affected by "fragmented" objects (i.e., objects that had missing cubes). Supporting the piecemeal strategy, researchers have observed task decomposition and rule-based learning strategies akin to a piecemeal strategy in complex tasks like inferring the motion of a complex pulley mechanism, gear movement and comparing 3D structures of molecules (Hegarty, 1992; Schwartz & Black,1996; Stieff, 2007) .

One way of determining what strategies people employ is to simply ask them, using either self-reports or strategy questionnaires and checklists. However, verbal reports can be encoded in several different ways and are subject to investigator-bias. The approach also harkens back to "introspective" approaches of early psychologists, and might be best thought of as merely a starting point for research of an exploratory nature.

From our own experience with verbal reports, we find that very often participants either lack the awareness of available strategies or are unable to articulate their own strategy. For example, in response to the question "How did you decide if the two objects were a match or a non-match" (see Figure 1), participants have responded saying "I looked at them and decided" or "I just knew." The additional investigator-prompts that such answers elicit may influence participants' choice of words and may only succeed in confirming the investigators own biases. Further, self-reports can be an ineffective method in developmental research with young children who have yet to develop a vocabulary to comprehend and respond to questions on strategy flexibility. Similarly, checklists or cognitive strategy/styles questionnaires are limited by their suggestive nature (Hegarty, 2010).

As an alternative, RT has been investigated as an indicator of strategy selection (Heil & Jansen-Osmann, 2008; Kail, et al., 1979; Kanamori & Yagi, 2002; Shepard & Metzler, 1971) with the idea that differences in RT reflect fundamental differences in strategy selection (Heil & Jansen-Osmann, 2008; Kail, et al., 1979). For example, a piecemeal strategy by definition requires multiple comparisons between the corresponding parts of a stimulus and, all other things being equal, would logically result in a longer RT with increasing stimulus components assuming encoding and rotation is completed for all parts of the stimulus before a decision is recorded. A holistic strategy, conversely, involves rotation along the horizontal and/or vertical axis and is independent of the number of constituent components of a stimulus. Therefore, a stable RT across varying numbers of components in a stimulus should be reflective of a holistic strategy. The biggest drawback of the RT approach is the inability to translate findings into effective spatial techniques to reduce individual or sex differences. Differences in slopes and intercepts when plotting RT on the y axis and angular disparity between objects of comparison on the x axis, are definitely suggestive of individual differences in mental transformation processes but fail to 'show' the researcher what the good performer is doing differently and at which stage of the mental transformation process. Knowledge of the RTs of good- and bad-rotators requires extrapolation to underlying cognitive strategies before it can be used to inform training techniques.

In addition to cognitive strategies, researchers have been interested in studying how participants flexibly employ these strategies. However, strategy flexibility is difficult to measure in real time. This is especially the case if participants rapidly alternate between strategies, which they may do in order to try to determine the best strategy. What is needed is a method that measures, in real time, the implementation of different strategies so that alternating between those strategies can also be measured. Strategy flexibility, which involves alternating between strategies based on task demand, although alluded to in several strategy selection studies as a future area of research (see Kozhevnikov, 2007), is difficult to study using these traditional measures of strategy selection.

To address some of the challenges faced using these traditional self- report and RT approaches and to supplement our knowledge of strategy selection, some researchers have adopted eye-tracking techniques to observe different strategies used during mental rotation (Just & Carpenter, 1976; Khooshabeh & Hegarty, 2010). With eye-tracking, researchers can track eye movements to measure the participant's path of attention to particular details of a visual stimulus. In other words, we capture those components of the object that the individual deems to be attention-worthy, and we also capture the order in which the different components capture the participant's attention (Duchowski, 2007). Eye fixations can be used to determine the amount of processing devoted to component stimulus features. The temporal sequence (or scanpath) is an important piece of information absent in self-report and RT approaches and provides information pertaining to comparison and confirmation strategy stages. Finally, eye-tracking techniques have been used effectively in developmental research with infants as young as 3-6 months (Aslin & McMurray, 2004; Gredebäck, Johnson & von Hofsten, 2009) and thus overcomes the challenge faced when studying strategy selection in young children with traditional methods.

There is precedent for the use of eye tracking in mental rotation literature. For example, Just and Carpenter (1976) used a corneal reflectance eye-tracking system to record eye movements to study the stages in cognitive processing during mental rotation. They found that total RT could be broken down into 3 stages –search, transformation and comparison, and confirmation. With the rapid advancement of eye-tracking methods (Nazareth, Odean & Pruden, 2017; Odean, Nazareth, & Pruden, 2015), other researchers have begun to use this approach too. For example, Khooshabeh and Hegarty (2010) used eye-tracking to compare the ratio of number of within-object fixations to the number of between-objects fixations. So in a holistic strategy, a participant would fixate once on the left object before transitioning to fixate once on the right object before recording their response and the ratio would be equal to 1. In comparison, a piecemeal strategy would require multiple fixations on one object before a transition is made to the next object for comparison and would consequently have a ratio greater than 1. The authors found that high-spatial individuals —as measured by a battery of spatial tests —tended to use a holistic strategy (ratio = 1) and low-spatial individuals used a piecemeal strategy (ratio >1). However, two different eye movements can be mistakenly encoded as a single strategy based on the within-object/between-objects ratio (see Figure 2). The "ratio" approach to eye movement analysis loses information about the duration and sequence of events (scanpath). The number of fixations, hereafter referred to as fixation count, is but one characteristic of eye movements. Table A2 describes the different ways in which eye movement data can be parsed to provide different types of information about participant eye-patterns. Eye-trackers are a rich source of eye-pattern data but unless they are accompanied by powerful pattern-detecting statistical analyses, we may be limiting its applicability to cognitive research.

Although limited in spatial strategy research, eye tracking has been used extensively in language-learning, scene perception, reading and math strategy research (Fernald, Pinto, Swingley, Weinbergy & McRoberts, 1998; Henderson & Hollingworth, 1999; Huettig & Altmann, 2005; Huettig, Rommers & Meyer, 2011; Ischebeck, Weilharter & Korner, 2016; Nation, Marshall & Altmann, 2003; Pyykkönen, Matthews & Järvikivid, 2010; Rayner, 1998; Rayner, Binder, Ashby & Pollatsek, 2001; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy,1995; Trueswell, Sekerina, Hill & Logrip, 1999). Although eye-tracking analyses in these areas of research include number of fixations on target items, time-to-first fixation, and even temporal analyses by breaking down fixation sequences between-items, these analyses do not capture within-subject, within-item mental transformations involved in spatial tasks.

This paper proposes a novel approach to analyzing eye movement data in the context of strategy selection and flexibility during mental rotation. With the help of a Tobii X60 eye-tracker, we record the eye movement parameters of fixation count, visit count, fixation duration and visit duration (see Table A2) during mental rotation on a standard Shepard and Metzler (1971) mental rotation task. We advocate for the use of latent variable mixture models like latent profile analysis to identify eye-patterns suggestive of strategy selection. In addition, we introduce the novel use of a changepoint detection algorithm with eye-tracking data to measure strategy flexibility.

We list below our main research questions and make predictions consistent with the literature on holistic and piecemeal strategy selection.

(1) What are the different cognitive strategies –indicated by eye-patterns – available to select from during mental rotation? We hypothesize that there will be two distinct eye-patterns indicative of strategy selection; a switching eye-pattern (i.e. high within- and between- object eye movement) akin to a piecemeal strategy and a fixating eye-pattern (i.e. low within- and between- object eye movement) akin to a holistic strategy. Our assumptions about these eye-movement parameters are based on definitions of holistic and piecemeal strategies (Heil & Jansen-Osmann, 2008; Khooshabeh, et al., 2013). We attribute high within- and between- eye movement to the piecemeal strategy since we expect participants to be looking at the distinct arms of one object (within-object) and making multiple comparisons with its paired object (between-object). We attribute low within- and between- eye movement to the holistic strategy since we expect participants to be encoding one object as a whole (within-object) and consequently making fewer comparisons with its paired object (between-object). (2) Which cognitive strategy is associated with better mental rotation performance? Consistent with previous eye-

tracking research (Khooshabeh, et al., 2013), we predict that participants demonstrating a fixating eye-pattern akin to a holistic strategy would perform significantly better than participants demonstrating a switching eye-pattern akin to a piecemeal strategy. (3) Are there sex differences in strategy selection? We predict sex differences in strategy selection with male participants being significantly represented in the eye-pattern associated with high mental rotation performance. (4) Is there a rigid adherence to one strategy or is there strategy flexibility? We predict that alternating between strategies (i.e., strategy flexibility) would be indicative of better mental rotation performance due to the ability to select an optimum strategy based on task complexity. (5) Are there sex differences in strategy flexibility? We predict that male participants would demonstrate significantly greater strategy flexibility than female participants. Research questions 1, 2 and 3 will be used to establish similarities between our eye-tracking data and previous approaches in distinguishing between two types of strategies during mental rotation. Research questions 4 and 5, however are designed to delve below the strategy dichotomies to investigate the less obvious role of strategy flexibility in mental rotation tasks.

Method

Participants

Adult university students enrolled in undergraduate psychology courses, between the ages of 18-25 were recruited from a university in the Southeastern United States. Undergraduate students could sign up on SONA SYSTEMS, a university-wide online recruitment website. They received one SONA credit for participation in the study that could be used as extra credit. The final sample consisted of 145 participants (M_{age}= 21.16

years, *SD*= 2.75 years; 48 male, 97 female). In the event of system crashes or technical issues with the eye-tracker, eye-tracking data were not analyzed if more than 50% of participant eye-movements were lost. Participants represented an ethnically (68% Hispanic/Spanish/Latino) and racially (62% White, 15% African American, 8% Asian and 20% other/mixed race) diverse group. The current research has received the university's Institutional Review Board approval (Protocol number: IRB-13-0367. Title: Blokus).

Materials

A standard Dell desktop with 2.80 GHz processor and 8 GB of RAM was used to present the stimuli. A Tobii X60 eye-tracker was used to record participant gaze data (see Table A1 for technical specifications of the eye-tracker). Stimuli were presented on a Sony Bravia (model number: KDL-46EX500) flat screen TV monitor using Tobii Studio 3.2, an eye-tracking analysis and visualization software compatible with Tobii X60 eye-trackers. A different standard Dell Desktop with 3.40 GHz processor and 6 GB of RAM were used to analyze the Tobii recordings. Data were analyzed using statistical software MPLUS v7, R, and SPSS v20.

Measure

Mental rotation task (MRT; Shepard & Metzler, 1971). The experimental stimuli were based on the Shepard and Metzler (1971) mental rotation task (MRT), which comprises pairs of objects of three dimensional, asymmetrical assemblies of cubes (see Figure 1). In each pair, the objects are either: (1) identical assemblies of cubes but rotated at different angles from each other (i.e., a match); or (2) assemblies of cubes that are mirror images of each other but rotated at different angles (i.e., a non-match). Participants were asked to decide if the two objects in each stimulus were identical or mirror images of each other. There were 20 unique pairs of match stimuli and 20 of non-match (i.e. mirror images) stimuli for a total of 40 stimuli. The presented orientation of the two objects i.e. the angle of rotation by which the two objects differed was random and ranged from 0 to 180 degrees (in 20 degree increaments) along the vertical axis. In order to counterbalance for practice effects and fatigue, the 40 stimuli were divided into two random sets of 20 stimuli each. The two sets of 20 stimuli were alternately presented at the beginning of a test condition. The two images were rotated along the same horizontal plane only.

In the *practice phase*, participants saw three solved trials where they were simply told what the answer was without solving the problem themselves. This was followed by three practice trials for which participants received feedback on whether they were correct or incorrect. In the *testing phase*, there were 40 trials with three rest periods of 30 seconds each after every 10 trials. There were no time constraints; once the participant recorded their response to a stimulus (i.e. match or non-match), the presentation automatically moved to the next stimulus. Responses were recorded using a computer mouse connected to the recording computer. The mental rotation score (MRT score) was a number ranging from 0 to 40 determined by the number of accurate responses (e.g., correctly indicating a match was a match or that a non-match was a non-match) on the MRT. This score was used to gauge a participant's mental rotation performance and was used as a dependent variable in the analysis for the current research study.

Procedure

Participants between the ages of 18 and 25 years were recruited from a

Southeastern University using the online recruitment website SONA SYSTEMS. Participants received course credit for the completion of the study. Participants were asked to carefully read and sign an adult consent form pre-approved by the Institutional Review Board (IRB) at the beginning of their laboratory visit. If a participant agreed to participate in the study after reading the study information on the consent form and listening to the study description by the trained research assistant, they were asked to enter the eye-tracking testing room. The eye-tracker room consisted of a desk-mounted Tobii X60 eye-tracker and a large screen TV monitor. The MRT was presented on the TV connected to the Tobii X60 eye-tracker. At the beginning of the MRT, participants were taken through a calibration process. The calibration process is an important first step in recording eye-movement. If the eye-tracker failed to calibrate, the process was repeated after moving the participant's position in front of the eye-tracker. Upon successful calibration, participants began the 40-item MRT. The eye-tracking component of the study took on average 20-25 minutes for completion from calibration at the beginning of the task to calibration at the end of the task. At the end of the eye-tracking component of the experiment, each participant was asked to fill out an online survey, which included basic demographic information such as participant sex, education level and SES. The online survey study component took on average 15-20 minutes for completion.

Rationale for Analyses

Strategy selection. In the present study, we used two analysis approaches to investigate strategy selection: (1) the available "ratio" approach from Khooshabeh and Hegarty (2010) and (2) a novel latent profile approach.

For the first analysis we employed the procedure from Khooshabeh and Hegarty, 2010,which computes strategy as the ratio of number of fixations within an object to the number of saccades i.e. switches made between two objects. For a holistic strategy, the number of within-object fixations is expected to be equal to the number of between-object saccades, resulting in a ratio close to 1. In contrast, for a piecemeal strategy, the number of within-object fixations is expected to be greater than the number of between-object saccades, resulting in a ratio significantly different and higher than 1.

For the second analysis of strategy selection, we employed a person-centered latent variable method to find patterns of eye-fixations based on four eye movement parameters —fixation count, fixation duration, visit count and visit duration (see Table A2). Latent profile analysis, a subset of latent variable mixture models (LVMM), is a flexible analytical technique that allows participant data to be grouped into patterns of similar behavior to determine the influence of these patterns on other variables of interest. In other words, it is a way of identifying sub groups of individuals sharing similar but directly unobservable characteristics (Berlin, Williams & Parra, 2014). Latent profile analysis is therefore an effective way of combining the different eye movement parameters to indicate eye-patterns based on count as well as duration of fixations and saccades. Latent profile analysis is a model-based technique which not only provides probability of class membership, but also goodness-of-fit statistics to compare models. One of the biggest advantages of latent profile analysis is its ability to deal with absolute magnitudes independent of sample size and the use of arbitrary score units of the latent indicators (Gibson, 1959). Using latent profile analysis, we aimed to answer our first three research questions – What are the different cognitive strategies –indicated by eye-patterns –available to select from during mental rotation? Which cognitive strategy is associated with better mental rotation performance? Are there sex differences in strategy selection?

Strategy flexibility. In addition to investigating strategy *selection*, we employed a changepoint analysis to explore strategy *flexibility*. We use a novel application of changepoint detection algorithms with eye movement data to investigate strategy flexibility. Changepoints are points in time where the statistical properties of the data prior to that time point are different from the statistical properties of the data after that time point. Any statistical property, e.g. mean, trend, variance, can be investigated depending on the structure of the data. For our data, we consider changes in mean and variance, more details on how we do this in described in the changepoint analysis section below. Unlike the "ratio" approach taken in previous eye movement data analysis (discussed in detail in the introduction section of the current paper), which ignores the sequence of fixations, changepoint detection can be applied to a time series to detect multiple changepoints in a single participant's eye-pattern over the course of the experiment, in this case, the MRT. Thus, changepoint detection moves the discussion away from switches in individual eye-fixations to switches in groups or patterns of eye-fixations. There are many different types of changepoint models, see Eckley, Fearnhead, Killick (2011) for a review. We consider within-subject fixations on each of the 40 stimuli and use changepoint analysis to determine if a person alternates between strategies during the MRT. We do not inform the algorithm when there is a change in stimulus and as such we can identify changes in strategy within a particular stimulus as well as between stimuli. Using changepoint analysis, we aimed to answer our last two research questions – Is there a rigid adherence to one strategy or is there strategy flexibility? Are there sex differences in strategy flexibility?

For both the latent profile and the changepoint analyses, an area-of-interest (AOI) was defined for each object of a stimulus. In other words, the object on the left formed AOI-1 and the object on the right formed AOI-2. This allowed us to identify when a participant switched between the left (AOI-1) and right (AOI-2) objects for comparison during mental rotation. In addition, we were able to count the number of fixations within AOI-1 and AOI-2. This allowed us to identify when a participant fixated on different features within the same AOI.

Results

Table 1 presents descriptive statistics for participant completion time, MRT performance, and eye-tracking parameters – fixation count, fixation duration, visit count and visit duration. Data were examined for normality and outliers. With regard to normality, skewness and kurtosis were used as indicators of univariate normality, with absolute values greater than 2.3 indicating severe non-normality (Lei & Lomax, 2005). Kurtosis and skewness for all variables were within acceptable ranges. Data were also evaluated for non-model based outliers by examining leverage indices for each individual and defining an outlier as a leverage score four times greater than the mean leverage. No outliers were found. To account for the unequal male and female sample sizes, we used bootstrapping with all gender-related mean differences tests and regression models (Efron & Tibshirani, 1986). Consistent with spatial literature (Nazareth, et al., 2013; Voyer, et al, 1995) males (M=32.33, SD=5.42) significantly outperformed females (M=28.36, SD=6.44) on the Shepard and Metzler (1971) MRT, t (143) = 3.67, p < .001, d = 0.67, bootstrap 95% C.I. on mean difference of 3.97 [1.96, 5.84].

Ratio Approach

We first calculated the strategy ratio (SR) as described in Khooshabeh & Hegarty (2010). SR was calculated as a ratio of the number of consecutive fixations within an object to the number of saccades (visits) made between the two objects of a stimulus. By the authors' definition, an SR value close to 1 would be indicative of a holistic strategy; an SR value greater than 1 would be indicative of a piecemeal strategy.

We found significant sex differences with males (*M*=2.35*, SD*=. 40) demonstrating a significantly lower SR than females (*M*=2.58*, SD*=. 81), *t* (143) = 2.28, *p* = .02, *d*=. 33, bootstrap 95% C.I. on mean difference of .23 [04, .44]. Khooshabeh and Hegarty (2010) did not report sex differences in SR value. However, this could be attributed to the fact that the authors did not find significant sex differences in mental rotation performance, unlike our sample and consequently may not have found strategy differences between male and female participants.

In addition, a one-sample t-test indicated that SR was significantly different from 1 (i.e. holistic strategy) for both male, *t* (47)= 23.64, *p*< .001, bootstrap 95% C.I. [2.25, 2.47] and female participants, *t* (96)=19.24, *p*<. 001, bootstrap 95% C.I. [2.43, 2.74]. This suggests that although male participants recorded lower ratios than female participants, both sexes in this sample appeared to use a piecemeal strategy as defined in Khooshabeh and Hegarty (2010) (see also, Just & Carpenter, 1976). This finding was corroborated by the non-significant correlation between SR and MRT score, *r* (143) = -.04, *p* =. 65, bootstrap 95% C.I. [-.20, .15]

We also calculated a two-stage hierarchical multiple regression with MRT score as the dependent variable. Participant sex was entered at stage one of the regression. SR was entered at stage two. The hierarchical multiple regression revealed that at stage one, participant sex contributed significantly to the regression model, F (1, 143) = 13.50, $p <$ .001, η^2 = .06, bootstrap 95% C.I. [1.98, 6.08] and accounted for 8.6% of the variation in MRT score. Introducing SR in the regression failed to explain any additional variation in MRT score and consequently the change in R^2 was not significant, F (2, 142) = .01, p=. 93, bootstrap 95% CI [1.88, 6.31].

To replicate previous research we used a median split as well as quartiles to divide our participants into high- and low- spatial participants based on their mental rotation performance. We ran a t-test to investigate if high- and low- spatial participants had significantly different SR values. For the median approach, there were no significant differences in strategies used by high-spatial participants (M=2.46, SD=. 41) and low-spatial participants (M=2.56, SD=. 01), t (102.96) = .88, p = .38, bootstrap 95% C.I. on mean difference of .10 [.37, .09]. A one-sample t-test indicated that SR was significantly different from 1 (i.e. holistic strategy) for both high-spatial participants, t (70)= 29.74, p< .001, bootstrap 95% C.I. [2.55, 2.36] and low-spatial participants, t (73)=14.80, p<. 001, bootstrap 95% C.I. [2.78, 2.37]. This suggests that although high-spatial participants recorded lower ratios than low-spatial participants –consistent with Khooshabeh and Hegarty, (2010) –both groups appeared to use a piecemeal strategy. We found similar results when using quartiles to categorize participants into high- and low- spatial groups and do not present those findings here for the sake of brevity.

This ratio analysis presents one potential method of using eye-movement data to investigate cognitive strategy use during mental rotation. However, it is limited in the information it provides because it fails to account for fixation duration i.e. how long a participant fixates at a point/arm of the object and visit duration, or stated simply, how long a participant stayed on an object using one or multiple fixations. As we discussed in the Introduction, the ambiguity involved in interpreting eye movement data cannot be resolved using the ratio approach. In addition, the ratio approach does not speak to another important aspect of spatial cognition, strategy flexibility. In the next sections, we present novel analyses that, we argue, provide a more sensitive measure of strategy use. These are latent profile and change point analyses. We present these findings in the order of our hypotheses outlined in the Introduction.

What are the different cognitive strategies – indicated by eye-patterns – available to select from during mental rotation?

To investigate strategy selection, we used the following eye movement parameters as latent predictors of strategy selection: fixation count, fixation duration, visit duration and visit count (see Table A2 for a detailed description of each eye movement parameter). Latent classes probabilistically determine participant membership (i.e., each participant has a unique probability for each of the latent classes, for a sum total of 1 across all latent classes; Lanza & Cooper, 2016; Muthén & Asparouhov, 2003). Although there are many different latent models to choose from, the current analyses used data that were cross-sectional and the latent class eye movement indicators (fixation duration, fixation count, visit duration and visit count) were continuous. Therefore, latent profile analysis was selected as the most appropriate latent variable mixture model (Collins & Lanza, 2013).

A total of 4 models (Class 1 to Class 4) were compared. The results of the model fit indices are presented in Table 2. The best fitting model was selected using information criteria (IC)-based fit indices – Bayesian Information Criteria (BIC; Schwarz, 1978), Akaike Information Criteria (AIC; Akaike, 1987) and Adjusted BIC (ABIC, Sclove, 1987). A lower value on the AIC, BIC and ABIC indicate a good model fit. Generally, the BIC is preferred for statistical model comparisons (Nylund, Asparouhov, & Muthen, 2007). In addition, entropy was used to assess the accuracy with which latent models classified individual participants into their most likely latent class. Entropy values range from 0 to 1, with 1 representing greatest classification accuracy. Finally, the Lo-Mendell-Rubin test (LMR; Lo, Mendell, & Rubin, 2001) and the Bootstrap Likelihood Ratio test (BLRT; McLachlan & Peel, 2000) were used to compare the model improvement between neighboring class models. These tests provide a *p* value that allows the determination of statistically significant improvement in model fit for the inclusion of one more latent class.

Based on the selection criteria described above, we selected the 2-class model as the best fitting model. The entropy value for the 2-class model was .86, signifying good precision in classification. High posterior probabilities were observed for the 2-class model with minimal (near zero fractional values) overlaps between classes (see Table 3) indicating high certainty of classification. The mean posterior probability for a student that belonged to latent class 1 in the 2-class latent profile model was .97. The mean posterior probability for a student that belonged to latent class 2 in a 2-class latent profile model was .93. We did not select the 3-class or 4-class models because their entropy values were almost identical to the 2-class model, suggesting potentially unimportant subclasses. The LMR p-value indicated that the 3-class and 4-class models did not fit the data significantly better than the more parsimonious 2-class model.

Table 4 presents the means of the individual eye movement latent indicator for the 2-class latent profiles. Latent class 1 — a fixating eye-pattern reflects use of a *holistic strategy* –had low mean fixation and visit counts (fewer switches between- and within-objects) in comparison to latent class 2 — a switching eye-pattern reflects use of a *piecemeal* strategy (greater number of switches between and within objects). The eye-tracking parameters involving duration (i.e., fixation duration and visit duration) were almost identical indicating a similar mean completion time between the two latent profiles.

In summary, for the latent profile analysis, we found that the parsimonious 2-class model best fit the eye movement data. This suggests that the two strategies of holistic and piecemeal in the mental rotation literature are reflected in differences in eye-patterns during the MRT, with the fixating eye-pattern mapping to the holistic strategy and the switching eye-pattern mapping to the piecemeal strategy. In all further analyses, we refer to latent class 1 as the fixating eye-pattern mapping to the holistic strategy and latent class 2 as the switching eye-pattern mapping to the piecemeal strategy.

Which cognitive strategy is associated with better mental rotation performance?

There was a significant difference in MRT score between latent class 1, fixating eye-pattern (i.e., holistic strategy; $M = 27.43$, $SD = 6.31$) and latent class 2, switching eye-pattern (i.e., piecemeal strategy; $M = 34.72$, $SD = 3.44$), $t\ (147.29) = 9.44$, $p < .001$; $d = 1.43$, bootstrap 95% C.I. on mean difference of 7.29 [8.70, 5.60] (see Figure 3), with those participants choosing a piecemeal strategy as defined by latent classes performing significantly better than those using a holistic strategy on the MRT.

Are there sex differences in strategy selection?

There was no significant association between the latent classes (i.e., holistic; piecemeal) and participant sex (i.e., male; female), χ^2 (1) =1.43, p = 0.23 (see Table 5). A two-stage hierarchical multiple regression was also calculated with MRT score as the dependent variable. Participant sex was entered at stage one of the regression given consistent sex differences seen in mental rotation tasks. Latent class membership was entered at stage two. Collinearity statistics (i.e., Tolerance and VIF) were all within accepted limits and there were no significant correlations between the independent variables of participant sex and latent class membership. Thus, the assumption of multicollinearity was met. The hierarchical multiple regression revealed that at stage one, participant sex contributed significantly to the regression model, F (1, 143) = 13.50, p < .001, η^2 = .06, bootstrap 95% C.I. [1.98, 6.08] and accounted for 8.6% of the variation in MRT score. Introducing the latent class variable explained an additional 22.6% of variation in MRT score and this change in R^2 was significant, F (2, 142) = 32.28, p < .001. Together, participant sex and latent classes accounted for 31.3% of the variance in MRT score, bootstrap 95% C.I. [1.73, 5.44]. Finally, we ran an additional regression model including the interaction between participant sex and latent class membership as a unique predictor of MRT score. In this model, both participant sex, F (1, 141) = 17.85, p < .001 and latent class membership, F (1, 141) = 46.81, p < .001 significantly predicted MRT score. However, the interaction between participant sex and latent class membership was not a significant predictor, F (1, 141) = 1.20, p = 0.27 η^2 = .01, bootstrap 95% C.I. [1.98, 7.34] suggesting that the type of strategy selected did not differentially affect male and female participants. The significantly higher variance explained by strategy selection and the lack of a significant interaction between strategy selection and participant sex suggests that both men and women should benefit equally from strategy training.

Strategy Flexibility

Changepoint analysis. In order to investigate strategy flexibility, we needed to identify points of change in eye-patterns from fixating–to–switching and vice versa. For example, let's say participant X used a fixating-pattern to mentally rotate the first 20 stimuli, used a switching-pattern at stimulus 21 and finally, went back to using a fixating-pattern at stimulus 35. Then, participant X would have alternated between strategies a total of 2 times during the 40-stimuli experiment. How can this type of pattern be detected? We argue that a changepoint detection algorithm with eye movement data can be used to investigate strategy flexibility.

From the original raw eye-tracking data the first step was to determine at which object the participant was looking. As the two objects were well separated we explicitly identified two clusters of data points. We used the x-coordinate (i.e. horizontal screen position of eye-fixations on left-right object) of the fixation points identified by the eye-tracker and clustered the data using K-means clustering (Hartigan and Wong, 1979; see Figure 4). Using the clustered data we then calculated a fixation proportion: the number of fixations before alternating to a different cluster-group (i.e. fixations on one object before switching to the second object), divided by the total number of fixations on each stimulus. We divided by the total number of fixations on a stimulus to control for the fact that some stimuli may have more fixations than others. The average number of fixations per participant was 395 (median 357) with the largest being 970 and 6 participants having less than 50 fixations.

We applied changepoint detection techniques on the fixation proportion to ascertain if a participant's strategy changes over time. To do this we used the *cpt.meanvar* function in the *changepoint* package (Killick & Eckley, 2014; Killick, Haynes & Eckley, 2016) in the *R* statistical program (R Core Team, 2016). This function uses the PELT algorithm (Killick, Fearnhead & Eckley, 2012) for fast and exact detection of potentially multiple changes in both the mean and variance of a time series. The default MBIC information criterion was used to determine if a change was present in the data. As we do not expect participants to be constantly switching strategies from one fixation to the next, we set the minimum number of fixations between changes to be the smaller of 10 and 20% of the data. The output of this function gave us both the changepoint locations as well as estimates of the mean and variance for each segment (strategy between switches). A resulting plot for a single participant with the means overlaid is given in Figure 5.

Is there a rigid adherence to one strategy or is there strategy flexibility?

On average, participants alternated between the fixating- and switching-eye-patterns 4 times during the 40-stimuli MRT (min = 0; max = 13, S*D* = 3.41). The large variability suggests that the number of changepoints could be a possible contributor to individual differences in MRT score. There was a significant correlation between MRT score and the number of changepoints detected, $r = .50$, $p < .001$, bootstrap 95% CI [.39, .60]. This finding indicates that participants who alternated between strategies more frequently performed better at the MRT than their less-flexible counterparts.

We conducted a second round of within-subject analyses to identify when the change occurred and its relation to performance immediately following the change in cognitive strategy. We found that on average, a participant recorded the correct response for 74% of stimuli during which a change in strategy occurred (guessing on the stimulus would be at 50%); the correct-response rate dropped to 72% for a stimulus immediately following the change in strategy. However, since we did not systematically manipulate the order of stimuli presented by angles of deviation between the two images, it is difficult to delineate the effects of strategy change from that of stimulus type.

Are there sex differences in strategy flexibility?

The number of changepoints detected for male participants (M = 5.03, SD = 3.70) was significantly higher than that for female participants (M = 3.49, SD = 3.22), t (130) = 2.35, p = .02, d = 0.44, bootstrap 95% CI for mean difference of 1.54 [.93, 6.68]. A two-stage hierarchical multiple regression was calculated with MRT score as the dependent variable. In model 1, participant sex was the only predictor. Model 2 included latent class membership (i.e., holistic; piecemeal) and number of changepoints detected (i.e., strategy flexibility). Collinearity statistics (i.e., Tolerance and VIF) were all within accepted limits. Thus, the assumption of multicollinearity was met. The hierarchical multiple regression revealed that at stage one, participant sex contributed significantly to the regression model, F (1, 129) = 11.24, p = .001, bootstrap 95% CI [1.31, 6.76] and accounted for 8.0% of the variation in MRT score. Strategy selection and strategy flexibility explained an additional 21% of variation in MRT score and this change in R^2 was significant, F (3, 127) = 17.00, p < .001. However, of the three model parameters, only participant sex (p = .01) and strategy flexibility (p < .001) significantly contributed to model 2. This is an important finding suggesting that strategy flexibility and not strategy selection may be responsible for individual differences in MRT score. The interaction between participant sex and strategy flexibility was not a significant predictor of MRT score, F (11, 105) = .60, p = .823, η^2 = .06, bootstrap 95% CI [-9.59, 2.14] suggesting that strategy flexibility equally benefited male and female participants during mental rotation.

Discussion

The current paper describes a novel application of the person-centered latent variable method to eye movement data in order to investigate individual differences in strategy selection (i.e., holistic; piecemeal) during mental rotation. Additionally, we introduce the use of a changepoint detection algorithm with eye movement data to investigate individual differences in strategy flexibility when solving a mental rotation task. We found two distinct eye-movement patterns during mental rotation –a fixating eye-pattern and a switching eye-pattern. The switching eye-pattern indicative of a piecemeal strategy was associated with high mental rotation performance. Thus, our latent profile analysis establishes similarities between our eye-tracking data and previous behavioral approaches in distinguishing between holistic and piecemeal strategies. In addition, strategy flexibility was found to be more important than strategy selection and this finding challenges the existing either-or approach to the study of strategy selection and strategy training in spatial research. Thus, unlike previous research we show that the traditional strategy selection approach to mental rotation is only half the story, and should be replaced by a discussion on strategy flexibility. Below, we discuss our findings in context to our original research questions and the possible interpretations and implications of these findings.

First, we asked what cognitive strategies are available to select from during a classic mental rotation task. This was an important step in order to demonstrate that strategy dichotomies traditionally discussed in the mental rotation literature could be detected in our eye-tracking data as well. Results of the previously used ratio approach to analyzing eye- movement data failed to demonstrate a relation between strategy selection and mental rotation performance. Although we found sex differences in the strategy ratio, men and women appeared to use a similar type of strategy i.e. a piecemeal strategy. Similarly, both high- and low- spatial participants appeared to use a piecemeal strategy. Thus, for our sample of participants the ratio approach failed to replicate theoretical and empirical research suggesting two distinct types of strategies during mental rotation. These findings support our argument that the ratio approach is less sensitive to the types of strategies used during mental rotation as it only considers the *count* of within-object fixations and between-object saccades and disregards the *duration* of these fixations and saccades. It also fails to provide information about the sequence of these fixations.

Contrary to the ratio approach, results of a latent profile analysis suggest two distinct eye-patterns when solving the MRT. The fixating eye-pattern had a low number of within-object and between-objects fixations, and the switching eye-pattern had a high number of within-object and between-objects fixations. This finding is consistent with strategy selection in the mental rotation literature, which describes a holistic strategy and piecemeal strategy. The fixating eye-pattern reflects use of a holistic strategy, which involves rotation of the complete object along a horizontal or vertical axis and therefore requires fewer comparisons within- and between-objects. The switching eye-pattern reflects use of a piecemeal strategy, which involves breaking up of an object into its constituent elements and consequently requiring frequent comparisons within- and between-objects. Our results thus provide additional support for idea that there are two distinct cognitive strategies for solving mental rotation tasks (Heil & Jansen-Osmann, 2008; Just & Varma, 2007; Kail, et al., 1979; Khooshabeh et al., 2013), and that these strategies map nicely to what has been dubbed in the literature as the holistic-piecemeal dichotomy.

Second, we asked which of these cognitive strategies are associated with better mental rotation performance. Recall that both Heil and Jansen-Osmann (2008) and Khooshabeh and colleagues (2013) found evidence that the holistic strategy use was associated with better performance on mental rotation tasks. Yet, others showed that the piecemeal strategy was useful in solving complex mental rotation tasks including inferring the motion of a complex pulley mechanism, gear movement and comparing 3D structures of molecules (Hegarty, 1992; Schwartz & Black, 1996; Stieff, 2007). These contradictory findings led us to examine which cognitive strategy was most effective in yielding higher MRT scores. Our results suggest that a switching eye-pattern reflective of the piecemeal cognitive strategy was associated with better mental rotation performance as compared to the fixating eye-pattern reflecting a holistic cognitive strategy.

A piecemeal strategy may comprise breaking up an object into constituent parts and systematically applying various rules and heuristics to attain a solution. This approach has been seen in spatial problem solving in STEM fields like engineering, chemistry and mathematics (Hegarty & Kozhevnikov, 1999; Schwartz & Black, 1996; Stieff, 2007; 2011; Stieff, et al., 2014) where novice students may initially use a holistic strategy, discover a rule and switch to a piecemeal strategy after extended practice. Hence, a person showing a switching eye-pattern reflective of a piecemeal strategy may have discovered a rule early on in the 40-stimuli task and was able to use it consistently to achieve a high MRT score.

We next asked whether there were sex differences in selection of a strategy to solve the MRT. There is prior work showing that males are more likely to use a holistic cognitive strategy, which in turn allows them to maintain a quick reaction time when solving mental rotation tasks even in the face of increasing stimuli complexity (Heil & Jansen-Osmann, 2008). We addressed potential sex differences by examining whether males and females simply differed in the cognitive strategy they used, but also looked at the contributions of participant sex, strategy selection, and the interaction between these two variables in predicting participant MRT scores. We found no significant sex differences in strategy selection. Thus, males did not consistently choose the holistic cognitive strategy any more than females. These results suggest that the male advantage in MRT scores may not necessarily be due to the use of one optimum strategy. What then might explain the reliable male advantage we see on the MRT? One potential explanation for this male advantage may be that males are *more flexible* in their strategy use, and not that they simply choose one strategy over the other. This was indicated by our changepoint analysis (discussed in more detail in the next section). Furthermore, while we found that participant sex and strategy selection each independently predicted MRT scores, the interaction between participant sex and strategy selection did not significantly predict MRT scores. This indicates that the type of strategy selected (i.e. holistic or piecemeal) does not differentially benefit/hurt males and females.

We next explored whether there is rigid adherence to one cognitive strategy or whether there is flexibility in cognitive strategy use when completing the MRT. To date, no studies have addressed real-time changes in strategies during multi-item mental rotation tasks and its relation to mental rotation performance. Using changepoint detection analysis, we found marked variability in strategy flexibility during the 40-item MRT, with some participants switching only once between strategies and others shifting more than a dozen times during the 40-time MRT. On average, participants switched strategies 4 times during the task. Critically, we also found that strategy flexibility significantly predicted MRT scores, even when controlling for participant sex and strategy selection. These findings suggest the existing presumption that participants choose an either-or strategy selection during mental rotation is not accurate. Indeed, our results show that participants who switch often between cognitive strategies do better on the MRT and strategy selection no longer significantly predicts mental rotation performance when controlling for strategy flexibility. Our findings are the first to explore the role of strategy flexibility on MRT.

The use of multiple strategies in response to varying task complexities has been observed in many domains including arithmetic (Kerkman & Siegler, 1993; Siegler, 1994; Siegler & Shipley, 1995) and problem solving (Crowley & Siegler, 1993). The study of strategy flexibility has been challenging to examine due to the difficulty in measuring flexibility using traditional RT and self-report methodologies. For example, another way strategy flexibility is studied is by using a between-group experimental design in which experimental groups receive different instructions and task complexities. While these studies are informative, as experimental manipulations they are less ecologically valid, and do not investigate strategy selection in real time, as we do here.

The last, and most important, research question we examined pertained to whether there are sex differences in strategy flexibility and whether these sex differences in strategy flexibility explain variability in MRT performance. Our results show a significant sex difference in strategy flexibility with male participants demonstrating higher strategy flexibility than female participants. Recall here that there were no sex differences in strategy selection. This finding suggests that males and females do not show a preference for holistic or piecemeal strategies; however, males show significantly higher strategy flexibility than females, which could explain the male-advantage in spatial tasks. This is an important finding with implications for spatial training interventions. There was no significant interaction between strategy flexibility and participant sex in predicting MRT score, indicating that strategy flexibility benefited both males and females. This suggests that spatial training interventions are likely to not only be beneficial for males, but for females as well.

Future Research

In the current study, we have demonstrated that strategy flexibility significantly predicts MRT score. Strategy selection, however, fails to explain additional variance in MRT score over that accounted for by participant sex and strategy flexibility. This is an important finding, suggesting that improvement in mental rotation cannot be attained merely through training in the use of one or the other strategy. Future research should examine how spatial experience and training can enhance strategy flexibility. Training interventions aimed at enhancing spatial skills need to design and assess the benefits of strategy flexibility training. Unlike traditional data collection methods and statistical analyses, the eye-patterns provide useful visual representations of the two strategies used during mental rotation and as such, can be used to define important stimuli features for spatial training. In addition, future research is needed to investigate the developmental trajectory of sex differences in strategy flexibility. One plausible explanation is that boys engage in more number of male sex-types spatial activities than females which may provide them with more opportunity to practice different strategies for problem-solving (Nazareth, et al., 2013, Newcombe, Bandura & Taylor, 1983).

The importance of strategy flexibility also highlights the fact that strategy selection is a function of the stimulus properties. However, what makes a participant alternate between strategies? Which stimuli properties determine an optimum strategy? An item-wise analysis of the different stimulus properties is needed to determine when a person chooses to use one of the other strategies. In the current research, the correct response following a change in strategy was higher than chance (guessing would be at 50%) does support our finding that strategy flexibility is associated with higher performance. However, there was no specific stimulus for which the majority of the participants switched strategies. We looked at the top five stimuli when change in strategy occurred but we failed to find any similarities between these stimuli (large angle of rotation, only mirror-image items, etc.). Since the items were counterbalanced, these 5 stimuli could have occurred at any time and were not always presented at the beginning or end of the task. Also, as stated earlier, we had not systematically manipulated the presentation of items in increasing angles of rotation. Hence, it is difficult to draw any conclusions from these five stimuli and consequently we left it out of the results section. Further eye-tracking research should investigate strategy-selection and flexibility at an item level through the careful manipulation angles of disparity between the paired items, presentation of matched and non-matched items and finally the order of presentation.

Limitations

In the current research we propose two novel approaches to analyzing eye-tracking data —the person-centered latent profile analysis and changepoint analysis. We suggest two eye-patterns indicative of holistic and piecemeal strategies. We also suggest that strategy flexibility and not strategy selection is predictive of mental rotation performance. These findings however are based on the Shepard and Metzler (1971) mental rotation task. Further research is needed in order to replicate and validate the two eye-patterns we proposed, across stimuli type and complexity. For our interpretations of the eye-patterns we are relying heavily on the theoretical framework created by and guiding subsequent verbal reports and RT spatial research. It is important to keep in mind that we are attempting to match eye-patterns with traditional descriptions of cognitive strategies. With increasing use of eye-tracking technology we expect that these existing strategy definitions will eventually evolve to create a new theoretical framework, hopefully focused on cognitive flexibility (as we propose here) as opposed to a strategy dichotomy, which widely exists today.

This research was designed to study individual differences in cognitive strategy selection and flexibility. Although we present our findings on sex differences, it is important to highlight the smaller proportion of males to females in the sample and a lack of significance could be attributed to a lack of power to detect a small sex difference. Given the unbalanced male-female sample size, our ability to draw firm conclusions about the link between participant sex and mental rotation skill was limited. Future research needs to be conducted with a larger sample consisting of equal number of males and female participants. It is also important to highlight that differences in scoring techniques of mental rotation tasks can result in different performance scores for men and women. For example, although men show a greater number of correct responses, they lose their advantage when a ratio of correct responses to items attempted is calculated (see Goldstein, Haldane & Mitchell, 1990).

Conclusion

Strategy selection and strategy flexibility are difficult to study using traditional data collection methods and statistical analyses. Our ability to enhance spatial thinking skills through experience and training, specifically for low-performing females depends heavily on our understanding of these underlying cognitive processes. We advocate for a shift in the traditional either-or strategy debate to investigating strategy flexibility using newer technologies and statistical algorithms. Although eye-tracking is not a novel approach by itself and is a rich source of information, conventional t-tests and median splits severely limit our interpretation of eye movement behavior. The statistical approaches to eye-tracking data that we employ in the current paper can be used beyond spatial research and is applicable to eye-tracking analyses in any domain.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317-332.

Aslin, R.N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. Infancy, 6, 155-163. Doi: 10.1207/s15327078in0602_1

Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, *39*(2), 174-187.

Butler, T., Imperato-McGinley, J., Pan, H., Voyer, D., Cordero, J., Zhu, Y. S., Stern, E., Silbersweig, D. (2006). Sex differences in mental rotation: Top-down versus bottom-up processing. *NeuroImage*, *32*(1), 445–456. http://doi.org/10.1016/j.neuroimage.2006.03.030

Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. *John Wiley & Sons*. http://doi.org/10.1002/9780470567333.ch2

Crowley, K., & Siegler, R. S. (1993). Flexible strategy use in young children's tic-tac-toe. Cognitive Science, 17, 531-561. doi: 10.1207/s15516709cog1704_3

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*(1), 25-32.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer Science & Business Media.

Eckley, I.A., Fearnhead, P. and Killick, R. (2011) Analysis of changepoint models. Chapter 10 in Bayesian Time Series Models, eds. D. Barber, A.T. Cemgil and S.

Chiappa, Cambridge University Press

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, 54-75.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228-231.

Ganesh, S., van Schie, H. T., Cross, E. S., de Lange, F. P., & Wigboldus, D. H. (2015). Disentangling neural processes of egocentric and allocentric mental spatial transformations using whole-body photos of self and other. *NeuroImage*, *116*, 30-39.

Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, *24*(3), 229-252.

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition*, *18*(5), 546-550.

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye tracking in infancy research. Developmental Neuropsychology, 35, 1-19. Doi: 10.1080/87565640903325758

Guillot, A., Champely, S., Batier, C., Thiriet, P., & Collet, C. (2007). Relationship between spatial abilities, mental rotation and functional anatomy learning. *Advances in Health Sciences Education*, *12*(4), 491-507.

Halpern, D. F.(1992). *Sex differences in cognitive abilities (2nd ed.).* Hillsdale, NJ:Erlbaum.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100-108.

Hegarty, M. (2010). Components of Spatial Intelligence. *The Psychology of Learning and Motivation*. http://doi.org/10.1016/S0079-7421(10)52007-3

Hegarty, M. (1992). Mental animation: inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1084.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of educational psychology*, *91*(4), 684.

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology (2006)*, *61*(5), 683–689. http://doi.org/10.1080/17470210701822967

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*(5), 438-443.

Hoyek, N., Collet, C., Rastello, O., Fargier, P., Thiriet, P., & Guillot, A. (2009). Enhancement of mental rotation abilities and its effect on anatomy learning. *Teaching and learning in medicine*, *21*(3), 201-206.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-B32.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, *137*(2), 151-171.

Ischebeck, A., Weilharter, M., & Körner, C. (2016). Eye movements reflect and shape strategies in fraction comparison. *The Quarterly Journal of Experimental Psychology*, *69*(4), 713-727.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441-480.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 153-191.

Kail, R., Carter, P., & Pellegrino, J. (1979). The locus of sex differences in spatial ability. *Perception & Psychophysics*, *26*(3), 182-186.

Kanamori, N., & Yagi, A. (2002). The difference between flipping strategy and spinning strategy in mental rotation. *Perception*, *31*(12), 1459–1466. http://doi.org/10.1068/p3325

Khooshabeh, P., & Hegarty, M. (2010, March). Representations of Shape during Mental Rotation. In *AAAI Spring Symposium: Cognitive Shape Processing*.

Khooshabeh, P., Hegarty, M., & Shipley, T. F. (2013). Individual differences in mental rotation: Piecemeal versus holistic processing. *Experimental Psychology*, *60*(3), 164–171. http://doi.org/10.1027/1618-3169/a000184

Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style. *Psychological bulletin*, *133*(3), 464.

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the Visualizer-Verbalizer Dimension: Evidence for Two Types of Visualizers. *Cognition and Instruction*, *20*(1), 47–77. http://doi.org/10.1207/S1532690XCI2001_3

Killick, R., & Eckley, I. (2014). changepoint: An R package for changepoint analysis.

Journal of Statistical Software, *58*(3), 1-19.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590-1598.

Killick R, Haynes K and Eckley IA (2016). changepoint: An R package for changepoint analysis. R package version 2.2.2, <URL: http://CRAN.R-project.org/package=changepoint>.

Kerkman, D. D., & Siegler, R. S. (1993). Individual differences and adaptive flexibility in lower-income children's strategy choices. Learning and Individual Differences, 5, 113-136. doi: 10.1016/1041- 6080(93)90008-G

Lanza, S. T., & Cooper, B. R. (2016). Latent Class Analysis for Developmental Research. *Child Development Perspectives*, *10*(1), 59–64. http://doi.org/10.1111/cdep.12163

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural equation modeling*, *12*(1), 1-27.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*(4), 344-351.

McLaughlin, G., & Peel, D. (2000). Finite mixture models.

Meneghetti, C., Cardillo, R., Mammarella, I. C., Caviola, S., & Borella, E. (2016). The role of practice and strategy in mental rotation training: transfer and maintenance effects. *Psychological Research*. http://doi.org/10.1007/s00426-016-0749-2

Muthén, B. O., & Asparouhov, T. (2003). Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in Mplus. *Mplus Web Notes*, *1*(6), 1–9. Retrieved from http://statmodel2.com/download/webnotes/webnote6.pdf\npapers3://publication/uuid/AE7B5FFA-70B2-4B1D-81FE-ABB9C8AFC662

Nation, K., Marshall, C. M., & Altmann, G. T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314-329.

Nazareth, A., Herrera, A., & Pruden, S. M. (2013). Explaining sex differences in mental rotation: role of spatial activity experience. *Cognitive processing*, *14*(2), 201-204.

Nazareth, A., Odean, R., & Pruden, S. M. (2017). The Use of Eye-Tracking in Spatial Thinking Research (pp. 239-260). In C. Was., F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research*. Hershey, PA:IGI Global.

Newcombe, N., Bandura, M. M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. *Sex Roles*, *9*(3), 377-386.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, *14*(4), 535-569.

Odean, R., Nazareth, A., & Pruden, S. M. (2015). Novel methodology to examine cognitive and experiential factors in language development: combining eye-tracking and LENA technology. *Frontiers in psychology*, *6*.

Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rine- hart and Winston.

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*, *25*(1), 115-129.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research*, *41*(7), 943-954.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333-343.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive science*, *20*(4), 457-497.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science (New York, N.Y.)*, *171*(FEBRUARY), 701–703. http://doi.org/10.1126/science.171.3972.701

Shepard, S., & Metzler, D. (1988). Mental rotation: effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 3.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. Current Directions in Psychological Science, 3, 1-5. doi: 10.1111/1467-8721.ep10769817

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), Developing cognitive competence: New approaches to process modeling (pp. 31-76). Hillsdale, NJ: Erlbaum

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and instruction*, *17*(2), 219-234.

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. http://doi.org/10.1037/a0034823

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 1632-1634.

Ter Horst, A. C., Jongsma, M. L. A, Janssen, L. K., Van Lier, R., & Steenbergen, B. (2012). Different mental rotation strategies reflected in the rotation related negativity. *Psychophysiology*, *49*(4), 566-573.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89-134.

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. http://doi.org/10.1037/a0028446

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.

Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental rotation. *Psychological science*, 0956797615585002.Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception & Psychophysics*, *31*(3), 201-209.

APPENDIX A

Tobii X60 Eye-Tracker Specifications and Terminology

Table A1

Technical Specifications for Tobii X60 eye-trackers (reproduced from the Tobii X60 technical manual)

Technical Specifications	Tobii X60
Dimensions (lXbXh)	320mm X 163mm X 85mm
Data Rate	60 Hz
Accuracy	Typical 0.5 degrees
Spatial Resolution	Typical 0.2 degrees
Head Movement Error	Typical 0.2 degrees
Tracking Distance	50-80 cm
Top head-motion speed	25 cm/second

Table A2

Eye-tracking Terminology

	Definitions (Reproduced from the Tobii Studio 3.2 user manual)
Fixation	A mathematical and statistical process by which raw data is parsed into fixations based on the velocity (in visual degrees per second) of the directional shifts of the eye. If the velocity of the eye movement is below a certain threshold the samples are classified as part of a fixation.
Fixation Duration	Measures the duration of each individual fixation within an AOI.
Fixation Count	Measures the number of times the participant fixates on an AOI
Visit	The interval of time between the first fixation on AOI1 and the first fixation on AOI2.
Visit Duration	Measures the duration of each individual visit within an AOI
Visit Count	Measures the number of visits within an active AOI

Note: AOI (Area of Interest) for the purpose of our analyses completely circumscribed each object of the stimulus. In other words, each stimulus had two AOIs, one for each object.

TABLES

Table 1

Descriptive statistics for study variables

Variable Name	Mean (SD)	Skewness	Kurtosis
Completion Time	9.11 (3.30)	1.17	1.35
MRT Score	29.54 (6.52)	0.27	1.08
Fixation Duration	0.23 (0.06)	-0.48	0.55
Fixation Count	12.38 (6.30)	0.78	0.66
Visit Duration	0.75 (0.33)	0.19	0.38
Visit Count	4.98 (2.31)	0.72	0.41

Standard error (skewness) = .21; standard error (kurtosis) = .40

Table 2

Fit Indices for the Latent Profile Analysis Models of Strategy Selection during Mental Rotation

No. Of Classes	1	2	3	4
No. Of free parameters	12	15	22	29
LL	-1376.34	-636.45	-575.36	-513.53
AIC	2776.67	1302.90	1194.73	1085.06
BIC	2813.87	1349.21	1262.65	1174.60
ABIC	2775.88	1301.72	1193.00	1082.79
Entropy	NA[c]	.86	.87	.90
LMR (p)	NA[c]	.03[b]	.38	.24
BLTR (p)	NA[c]	.00	.00	.00

Note. LL = Log Likelihood; AIC = Akaike Information Criterion; ABIC = Adjusted BIC; LMR = Lo-Mendell-Rubin Adjusted Likelihood Ratio Test; BLRT = Bootstrap Likelihood Ratio Test. There were 4 indicators: fixation duration (FD), fixation count (FC), visit duration (VD) and visit count (VC).

[b] Best-fitting model according to that index.

[c] LMR and BLRT are not available for the one-class model.

Table 3

Average Latent Class Probabilities for Most Likely Class Membership

Latent Class Model	Latent Class membership			
	1	2	3	4
2-class				
Class 1	**.97**	.07	-	-
Class 2	.03	**.93**	-	-
3-class				
Class 1	**.94**	.0	.02	-
Class 2	.00	**.93**	.03	-
Class 3	.07	.07	**.95**	-
4-class				
Class 1	**.95**	.00	.05	.00
Class 2	.00	**.91**	.10	.00
Class 3	.02	.03	**.95**	.00
Class 4	.00	.01	.00	**1.00**

Table 4

Latent Class Means for Eye Movement Parameters in the 2-class Latent Model

Unstandardized means	Latent Classes 1	Latent Classes 2	Significance testing
Fixation Duration	.23 (.06)	.25 (.04)	t (143) = 2.02, p = .05
Fixation Count	9.41 (3.68)	20.47 (5.31)	t (143) = 14.45, p < .001
Visit Duration	.75 (.39)	.78 (.19)	t (143) = .49, p = .63
Visit Count	3.86 (1.29)	7.92 (1.61)	t (143) = 16.12, p < .001

*. Area-of-Interest (AOI) overlapped with each object of the mental rotation task

Table 5

Logistic Regression Coefficients for the 2-class Latent Profile Analysis with Mental Rotation Score and Participant Sex as Latent Class Predictors

Latent Predictors	Latent Class	
	1	2
MRT Score	-.232*	.232*
Sex	.048	-.048

FIGURES

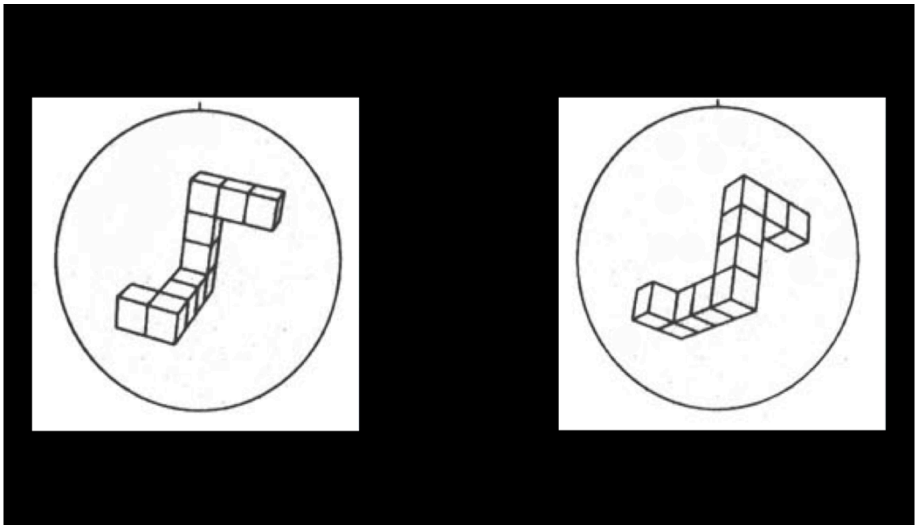

Figure 1. A sample stimulus used in the standard experimental condition, based on the Shepard and Metzler (1971) mental rotation task. The two objects are mirror images of each other and cannot be rotated into congruence. The correct participant response for this stimulus is ‘non-match’.

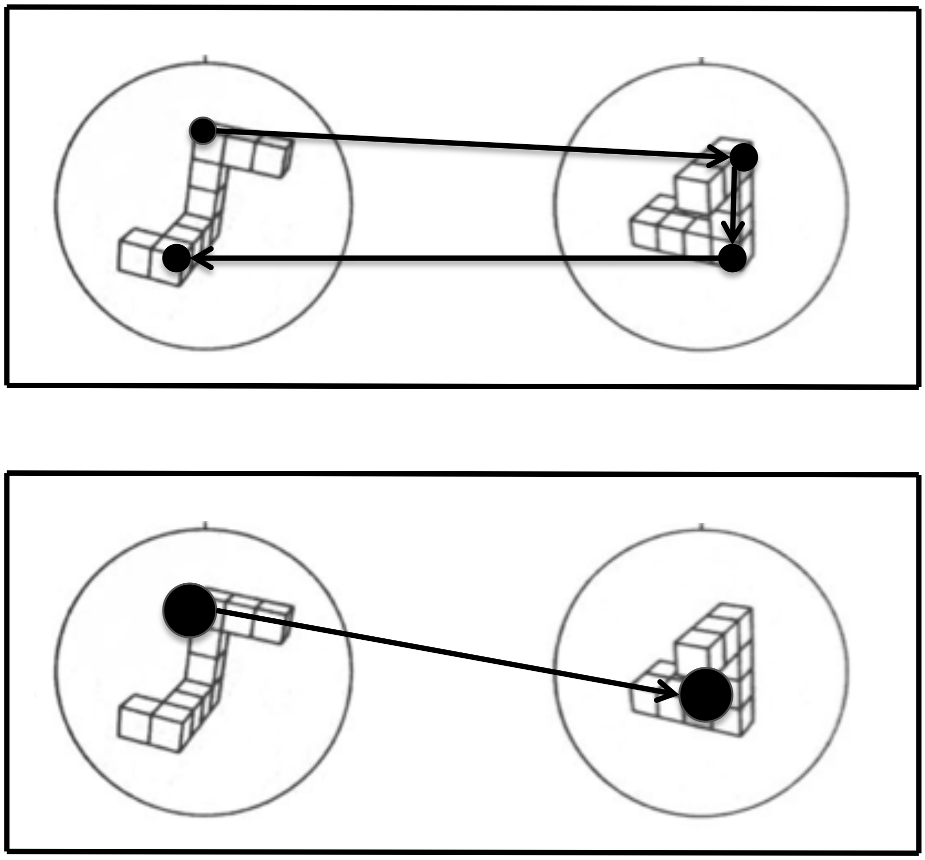

Figure 2. Top: within-object fixations = 2, between-object fixations = 2, ratio = 2/2 = 1. Completion time = 4 seconds. Bottom: within-object fixations = 1, between-object fixations = 1, ratio = 2/2 = 1. Completion time = 4 seconds.

Note: Dots indicate eye-position. Size of dot indicates duration of eye fixation.

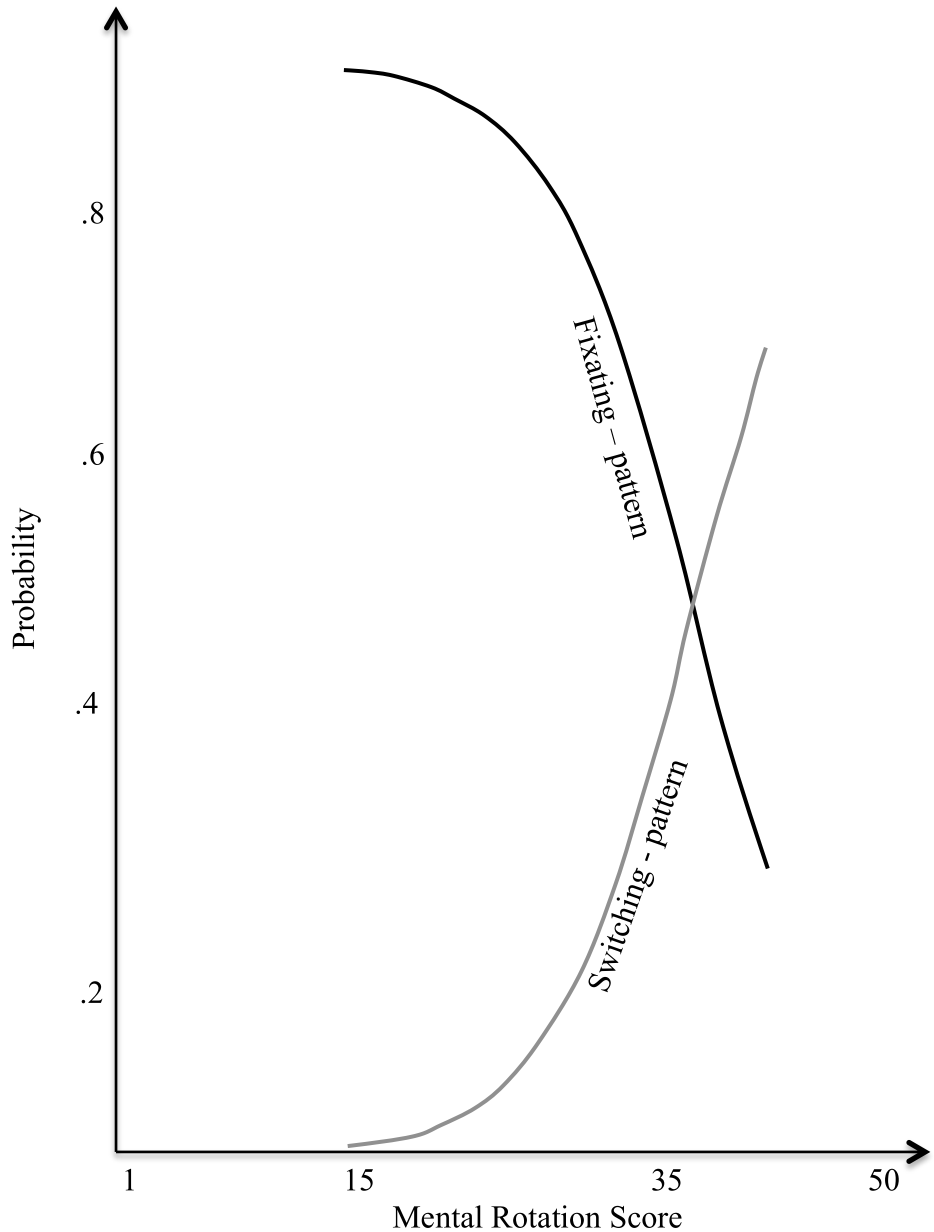

Figure 3. A 2-class latent profile model using mental rotation score (MRT) as a latent class predictor.

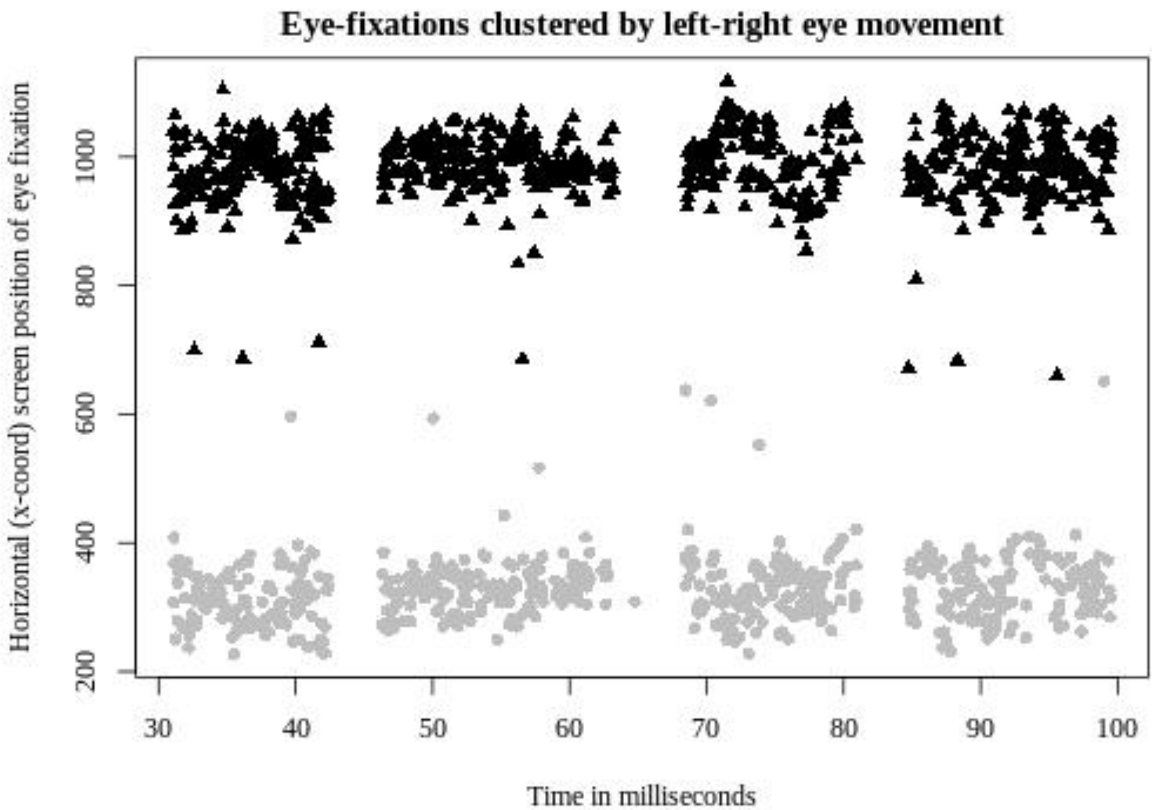

Figure 4. Cluster plot of participant eye movement. The y-axis represents the x coordinate (horizontal position) of a participant's eye-fixations on the television screen since the stimulus consisted of objects placed to the left and right of the screen. Thus, gray clusters indicate fixations on the left object and black clusters indicate fixations on the right object to be compared on the Shepard & Metzler (1978) mental rotation task.

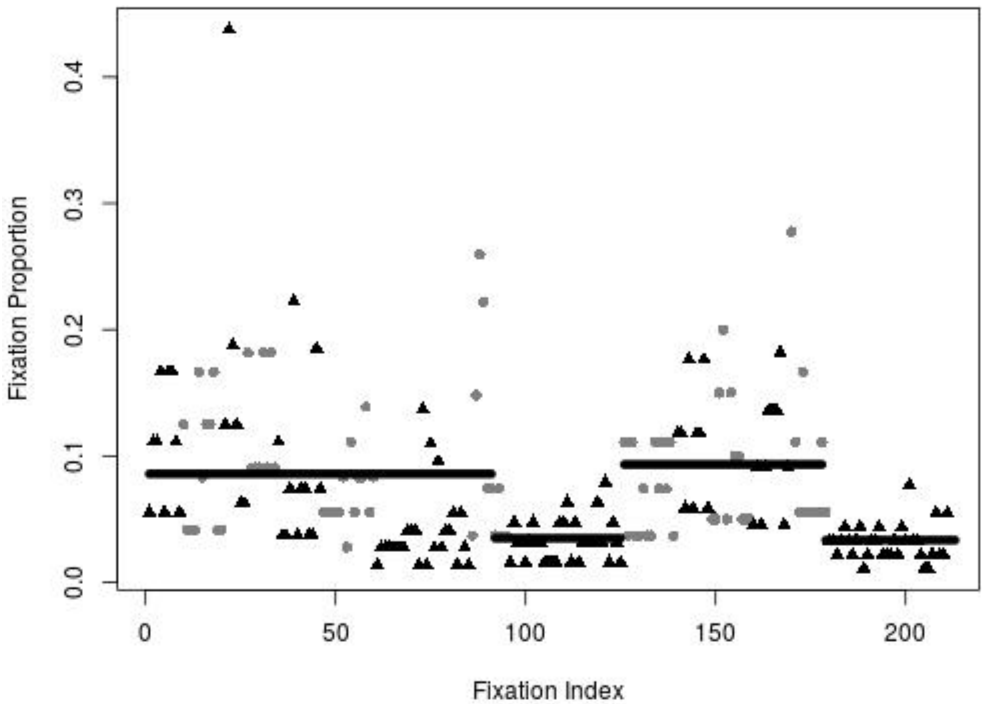

Figure 5. Changepoint detection using PELT. The above plot indicates the location of changes between strategies during the 40-item MRT (truncated at stimulus 15 for readability). The alternating grey and black data points indicate a change in the stimulus i.e. a participant starts looking at a new stimuli every time the data point switches colors. There are three change points detected for the current participant. The duration of the strategy selected is depicted by the overlaid black segments.