Using eye-tracking to understand the complex relations between attention and language in children's spatial skill development

by

Hilary E. Miller

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

Doctor of Philosophy
(Psychology)

at the
University of Wisconsin-Madison
2018

Date of oral examination: 5/16/18

This dissertation is approved by the following members of the Final Oral Committee:

Vanessa Simmering, Assistant Professor of Psychology
Martha Alibali, Vilas Distinguished Achievement Professor of Psychology
Jenny Saffran, Distinguished Professor of Psychology
Heather Kirkorian, Associate Professor of Human Development and Family Studies
Haley Vlach, Assistant Professor of Educational Psychology

Table of Contents

Abstract iii
Chapter 1: Introduction 1
  Theories of Language and Spatial Cognition 2
  The Role of Language in Spatial Development 5
    Effects of spatial words 5
    Beyond effects of spatial words. 8
  Visual Attention and Relation to Spatial and Problem Solving Tasks 12
    Eye-gaze patterns during tasks 13
    Manipulating eye-gaze patterns to improve performance. 16
Chapter 2: Current Study 18
Chapter 3: Method 24
  Participants 24
    Sample Size Rationale. 25
  Apparatus 25
  Design 28
  Procedure 31
    Recall tasks 31
    Production task 35
    Comprehension task. 35
    Calibration 36
  Method of Analysis 37
    Search scoring and exclusions. 39
    Offline calibration. 39
    Eye-gaze coding. 40
    Eye-gaze measure rationale. 45
    Reliability for eye-gaze data. 48
    Eye-gaze data quality analysis and exclusion. 49
Chapter 4: Results 56
  Effects of Cue Manipulations on Recall Accuracy 56
  Effects of Condition on Eye-Gaze 59
  Effects of Eye-Gaze and Condition on Recall Performance 63
Chapter 5: Discussion 68
  Effectiveness of Verbal versus Visual Conditions 74

- Children's Spontaneous Visual Attention 78
- Mechanisms of Developmental Change 82

References 87
Appendix A 100
Appendix B 103
Appendix C 106

Abstract

This dissertation investigated relations between visual attention and language as these factors relate to spatial skill development. It used eye-tracking to differentiate among hypotheses regarding whether: language causes changes in spatial cognition unrelated to changes in visual attention; language and visual attention both facilitate spatial performance but through different mechanisms; or language only relates to spatial performance in that it is a useful cue in directing visual attention? Four- to five-year-old children participated in two spatial recall tasks assessing their use of an intrinsic reference frame during recall (memory for relations among nearby objects). The task involved children finding a toy's location under a cup on an array of cups and landmarks after a 5 second delay and the array rotated. Children first participated in the baseline recall task receiving non-specific cues. Then children participated in the recall task again under conditions of verbal, visual, or no-specific cues (control). Consistent with past research, the results showed that the verbal cues were most effective in supporting spatial performance. Additionally the results showed that children's visual attention can be directed to support their spatial performance. Effects of verbal cues were partially mediated by children's visual attention; but verbal cues supported children's performance more than by simply directing their visual attention. This is the first known study to directly measure children's visual attention to predict their spatial performance. These results provide compelling support for a weak verbal encoding hypothesis that both language and visual attention support young children's spatial performance.

Chapter 1: Introduction

Spatial skills are basic cognitive processes essential for everyday behavior that also support higher level reasoning in Science, Technology, Engineering, and Math (STEM) fields (e.g., Vasilyeva & Lourenco, 2010; Wai, Lubinski, & Benbow, 2009). Spatial skills are important for everyday behaviors, including perceiving and remembering object locations, navigating and orienting in one's environment, reasoning about relations among objects, mentally rotating objects, and taking the perspective of others (Hegarty & Waller, 2005). Having strong spatial skills is important for success in STEM careers. For example, high school students' spatial skills predict performance in math and science classes and their later entry into STEM fields (Wai et al., 2009). Spatial skills support reasoning in STEM domains, such as through facilitating representations of molecule and atom structures in chemistry (H. Wu & Shah, 2004), making and using maps in geoscience (Kastens & Ishikawa, 2006), and graphic designing in engineering (Sorby, 2009). These relations among spatial skills and STEM learning emerge early in development. Spatial perception in infancy predicts preschool children's math skills (Lauer & Lourenco, 2016) and spatial skills at the start of preschool predict children's math skills in kindergarten (Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017). Considering the importance of spatial skills for everyday behaviors and later life achievement, it is essential to understand mechanisms supporting spatial development early in childhood.

Spatial language and visual attention are both candidate processes supporting spatial skill development (e.g., Miller & Simmering, 2018; Pruden, Levine, & Huttenlocher, 2011; Shusterman & Spelke, 2005). Both children's spatial language and visual attention skills are improving over development (e.g., Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Plude, Enns, & Brodeur, 1994), and likely interact to promote children's spatial development. Multiple studies have established relations between spatial language and spatial reasoning in adults and children (e.g., Loewenstein & Gentner, 2005; Pederson et al., 1998; Pyers, Shusterman, Senghas, Spelke, & Emmorey, 2010). Additionally, studies have demonstrated the importance of visual attentional processes in adult spatial cognition (e.g., Bailey, McNamara, Costello, Sridharan, & Grimm, 2012; Nazareth, Odean, & Pruden, 2017). However, there has been less research investigating the role of visual attention in supporting children's spatial skills. This limitation in understanding the role of visual attention in spatial development is partially due to research focusing only on outcome measures (i.e., number of correctly solved spatial problems) and not on process based measures (i.e., how children attend to and use various cues while solving spatial problems). There is thus a need to better understand the role of visual attention and how visual attention and language interact to support spatial development. This dissertation used eye-tracking as a process based measure to increase understanding of the types of spatial cues children attend to while performing spatial tasks and of whether and how visual attention and language interact to support spatial development.

Theories of Language and Spatial Cognition

Language is a tool that can augment cognition, with effects demonstrated in areas such as numerical representation (e.g., Miura, Kim, Chang, & Okamoto, 1988), categorization (e.g., Bowerman & Choi, 2003), analogical reasoning (e.g., Gentner, 2003), and spatial cognition (e.g., Landau, Dessalegn, & Goldberg, 2010). The role of language in spatial cognition has received much attention particularly due to various cultural groups having different spatial language experiences. For example, languages show differential preference for spatial reference frames (systems for representing object locations) as some languages such as English tend to code locations egocentrically (e.g., "left" and "right" of the body) while other languages such as Tseltal tend to code locations allocentrically (e.g., "north" and "south" cardinal positions, Levinson, 2003). Languages also code spatial categories differently as, for example, English separately categorizes containment and support relations while Korean separately categorizes tight fit and loose fit relations (Bowerman & Choi, 2003). Finally, some individuals have less knowledge of spatial words due to learning a recently developed language (Pyers et al., 2010) or due to not being exposed to a conventional language (Gentner, Özyürek, Gürcanli, & Goldin-Meadow, 2013). These cultural differences in language exposure lend themselves to testing the role of differential language experience on spatial cognition.

Primarily from investigations of cultural variability in language experience, two classes of theories have emerged for how language influences spatial cognition. One class of theories, which I will refer to as the strong verbal encoding hypothesis, proposes that language permanently alters spatial reasoning and without sufficient language one cannot reason to the same extent (e.g., Levinson, 2001). Some support for this hypothesis comes from studies having participants recall the position of three objects in a row on a table and having them move 180º and then see how the participants re-align the objects. This type of study typically shows that speakers of absolute reference frame languages will order objects relative to the exact position in the room while speakers of egocentric reference frame languages will order the objects in the same position relative to themselves (Pederson et al., 1998). Additionally, evidence comes from learners of Nicaraguan Sign Language (NSL), a recently developed sign language. Nicaraguans who were learning NSL when the language was first developing inconsistently used the spatial terms "left" and "right" compared to a later cohort of NSL learners. Pyers et al. (2010) tested the two cohort of NSL learners in a disorientation search task, in which the participants had to search for a hidden object in a corner of a rectangular room. In some conditions, one wall of the room was a distinct color and the task assessed whether participants could use the colored wall (a featural cue) to help them reorient in the room. Compared to the later cohort, the early cohort of NSL learners performed worse on the disorientation search task as they were unable to use the left/right relations of the colored wall to find the hidden objects. Similarly, in a non-cross-cultural comparison study using this task, adults who were engaging in verbal shadowing (disrupting their abilities to verbalize the task) performed worse in using the colored wall to reorient compared to control and non-verbal shadowing groups (Hermer-Vazquez, Spelke, & Katsnelson, 1999, but see Ratliff & Newcombe, 2008).

Another class of theories, which I will refer to as the weak verbal encoding hypothesis, proposes that language can temporarily augment cognition but does not permanently change our representations (e.g., Landau et al., 2010). Some evidence for this perspective comes from studies using similar reference frame tasks to Pederson et al. (1998). When task instructions are non-ambiguous regarding which reference frame to use, Tzeltal speakers were not impaired in using an egocentric reference frame, a reference frame that is incongruent with their native language (P. Li, Abarbanell, Gleitman, & Papafragou, 2011). Likewise, speakers of English and Korean can easily categorize spatial relations against their language dominant category when the relations are non-ambiguous and perceptually salient (Choi & Hattrup, 2012). These lines of research suggest that people can flexibly represent spatial relations even if their language does not include words for encoding those relations. Additional evidence comes from a non-cross-cultural study in which the researchers tested whether adults were spontaneously encoding spatial relations into language in a task where participants needed to remember two objects in ambiguous spatial relations. Feist and Gentner (2007) showed that only participants who received a linguistic label encoded the spatial relations into language. There was no evidence of spontaneous verbalizations for participants who did not receive labels. Across these studies, within both the strong and weak verbal encoding hypothesis, there is strong evidence that language can support spatial cognition and can guide people to reason in particular ways. However, the extent to which language permanently causes changes in spatial cognition and the extent to which people automatically code information into language is still a source of debate and a question beyond the scope of this dissertation.

The Role of Language in Spatial Development

In addition to work with adults, multiple studies have found effects of language on children's spatial performance. Effects of language on spatial performance have been shown across a diverse range of spatial tasks, including those assessing reference frame selection in spatial recall (Miller, Patterson, & Simmering, 2016); reorientation (Hermer-Vazquez, Moffet, & Munkholm, 2001; Shusterman, Lee, & Spelke, 2011); mental rotation and translation (Pruden et al., 2011); relational reasoning (Loewenstein & Gentner, 2005; Pruden et al., 2011; Simms & Gentner, 2008); and feature binding (Dessalegn & Landau, 2008, 2013; Farran & O'Leary, 2016). Most research on this topic has focused on the effects of spatial words on spatial performance with less attention given to children's general, not necessarily spatial, use of task-relevant language and to other non-linguistic processes that might underlie the effects found with language early in development.

Effects of spatial words

There are two main lines of research investigating the role of spatial words on children's spatial performance. One line of studies uses correlational methods and has found relations between children's spatial word use and their spatial performance. For example, studies have found that children who are better at producing spatial words relevant for a task, such as "left"/"right" (Hermer-Vazquez et al., 2001), "by"/"next to" (Miller et al., 2016), or "middle" (Simms & Gentner, 2008), tend to perform better on spatial tasks requiring encoding of such spatial relations.

Additionally, children whose caregivers produce more spatial words in free play settings tend to also produce more spatial words themselves, which in turn, predicts their later spatial performance (Pruden et al., 2011).

A second line of studies shows that providing children with spatial words that are relevant for a task during or before the task can enhance children's spatial performance (e.g., Dessalegn & Landau, 2008; Miller et al., 2016). For example, providing children with task-relevant verbal cues (e.g., "the yellow is on the left") improved 4-year-old children's performance in a feature binding task where children needed to remember the left/right relations among two sides of a colored square. These types of effects of providing children with spatial language during a task are fairly robust as they can last even two days after the initial language exposure (Loewenstein & Gentner, 2005). However, there tends to be a small window in which these benefits of providing language support are present (Dessalegn & Landau, 2013; Loewenstein & Gentner, 2005). Three-year-old children tend not to benefit from spatial language support, possibly because they have limited knowledge of the spatial words and tend not to use such words to support their performance. Six-year-old children also tend not to benefit from language support, presumably because they are already spontaneously encoding these cues into language.

Some studies using this paradigm of providing children with spatial words during a task have also suggested that these effects are unique to language as providing children with visual cues that are intended to be analogous to the verbal cues are not as effective in promoting spatial performance (Dessalegn & Landau, 2008; Miller et al., 2016). For example, in the feature binding task, children showed no improvements relative to a control group when provided with visual cues drawing attention to one side of the colored square (i.e., one colored side of the square flashed on the screen). Additionally, Miller et al. used a spatial reference frame recall task testing children's use of multiple reference frames during recall in a task where children needed to find a toy hidden under a cup on an array of multiple cups and landmarks. Specifically, they assessed whether children were recalling object locations relative to an egocentric (relative to the body), room-centered (relative to the room) and/or intrinsic reference frame (relative to other nearby objects). Similar to other studies finding effects of language, the authors found that verbal cues specifying relations among landmarks on the array (“the toy is hidden by the frog”) facilitated children’s performance relative to a control group. In their visual analog, Miller and colleagues visually drew children’s attention to the relation between the cups and landmarks by moving one of the nearby landmark(s) on top of the target hiding cup and then placing the landmark back in its original position. They found modest improvements with visual cues: children performed better relative to the control group, but worse than those provided with relevant language cues. Additionally, in an analysis of the trials that required children to use an intrinsic reference frame (trials that most highly depended on encoding relations among landmarks and cups), children’s performance in the visual condition was not significantly higher than in the control condition. Based on these outcome measures, this work suggests that language might be unique in promoting spatial performance, providing children with a qualitatively different way of reasoning about spatial relations, as it was the only factor that most strongly improved children’s performance.

Together these two lines of work (correlation and experimental) showing effects of spatial language on spatial performance demonstrate that spatial words are useful in facilitating children’s encoding of spatial cues. Based on this work, multiple researchers have theorized that children’s acquisition of spatial words facilitates their encoding of spatial cues, in turn enhancing their performance (e.g., Hermer-Vazquez et al., 2001; Loewenstein & Gentner, 2005). This perspective is consistent with either verbal encoding hypothesis on the role of language in spatial cognition from the adult literature as it suggests that the acquisition of spatial language can change or augment spatial reasoning.

Beyond effects of spatial words.

More recent work has investigated effects of non-spatial but task-relevant language on children's spatial performance, showing that spatial words are not the only type of language that can improve children's performance. For example, Dessalegn and Landau (2013) tested 4-year-olds in the feature binding task (described above). Children who heard language highlighting the asymmetric relation between the two colored halves (e.g., "the black one is prettier") showed better performance relative to a control group. Additionally, children who heard non-relational language that highlighted task-relevant cues also performed better than the control group. Using the disorientation search task (described above), Shusterman et al. (2011) showed that children improved their performance relative to a control group when they were exposed to non-spatial language that highlighted the utility of important spatial features, specifically a colored wall (e.g., "the red wall can help you find the sticker"). These effects showing benefits from non-spatial language are specific to task-relevant language; other types of language that do not draw attention to the important asymmetric relations or the utility of the cues (e.g., "the yellow is touching the black" and "look at this pretty wall") are not helpful (Dessalegn & Landau, 2013; Shusterman et al., 2011). Thus, the relevance of language seems to be the critical factor in supporting spatial performance.

However, this research has led to questions on the extent to which effects found with language are reflective of changes in children's verbal encoding or of more basic changes in children's visual attention skills. Despite evidence that task-relevant language facilitates spatial performance (e.g., Shusterman et al., 2011), this research is limited in accounting for how children might use this type of task-relevant language on their own to support their spatial performance when the language is not provided by an adult. If knowledge of relevant language is driving children’s spatial skill development, then once children can produce relevant words they should encode relevant relations and, in turn, perform better on spatial tasks. However, young children may know words that could be relevant for a task (as in the “prettier” example above) but may not use these words to facilitate their performance. For example, Farran and O’Leary (2016) used the feature binding task and found that only children who showed knowledge (productive and receptive) of particular spatial terms (e.g., “left” and “right”) increased their performance relative to baseline when an experimenter provided them with such terms. These results suggest that children who knew the relevant spatial terms only used them with experimenter support. Thus, language knowledge alone may not be sufficient to support performance; children must also use their knowledge effectively in the task. This raises important questions of what other mechanisms are needed to explain effects found with language on spatial development, as having language knowledge does not account for using such language.

Miller, Vlach, and Simmering (2017) investigated potential mechanisms relating children’s language and spatial performance. They tested whether children’s spatial skills were predicted more strongly by their production of spatial words versus their use of task-relevant language. Four-year-olds completed a spatial scene description task that assessed their adaptive use of task-relevant language and children were also tested on a set of spatial tasks and a receptive vocabulary task. Results showed that children who used language more adaptively (i.e., provided more relevant than irrelevant cues) in the spatial scene description task performed better on the spatial tasks, even when controlling for demographic and language factors previously shown to relate to spatial performance. These results provided evidence that producing spatial words is not enough to facilitate spatial performance; instead children's abilities in using task-relevant language was a stronger predictor of spatial performance.

Miller & Simmering (2018) took the research by Miller et al. (2017) a step further and asked whether the effect of children's adaptive language use resulted from children's abilities to use language or rather from more basic processes of attention to task-relevant information. They designed a memory task version of the spatial scene description task and found that children's memory for task-relevant cues predicted performance above and beyond how adaptive children were at using language and the other factors included in the Miller et al.'s study. In a second experiment, Miller and Simmering reduced demands on memory and production to confirm that children's performance in the memory and description tasks were primarily due to limitations in attending to task-relevant information and not from task-specific memory and production demands. This work suggested that developmental changes in attention to task-relevant information was more strongly predictive of children's spatial development than their abilities to use verbal encoding. In fact, it is important to consider that during the preschool years there are significant changes in children's abilities to selectively attend to relevant stimuli (e.g., Colombo & Cheatham, 2006; Plude et al., 1994), which is likely directly influencing how children approach and perform spatial tasks.

Other research has also found limitations in theories claiming that verbal encoding is a central mechanism underlying spatial skill development. For example, studies using the disorientation search task (described above) have shown performance improvements when featural cues (i.e., colored wall) in the testing room are more salient or stable. In these studies when the featural cues are more stable and salient, 2-year-old children can use these cues to reorient and solve the task, an age before they can use the words "left" and "right" for encoding (Learmonth, Nadel, & Newcombe, 2002; Learmonth, Newcombe, & Huttenlocher, 2001). Additionally, Nardini, Burgess, Breckenridge, and Atkinson (2006) tested 3- to 6-year-old children in a reference frame recall task (similar to the task used by Miller et al., 2016). On one trial during the task, after the child turned around for the delay, the experimenter asked the child "can you tell me where the toy is hiding?" The authors found that only 6-year-olds were able to spontaneously verbally recode the location of the toy relative to the landmarks on this "surprise" trial. However, 5-year-old children still successfully used the relative position of landmarks to recall the target location, suggesting that language is not critical to performing this task. Similar effects have been found in non-spatial working memory tasks whereby 5-year-old children tend to depend more on visual working memory to solve object recall tasks, while 10-year-old children tend to rely more on verbal working memory (Hitch, Halliday, Schaafstal, & Schraagen, 1988). Together these results show that children are not necessarily using verbal encoding during spatial tasks on their own to support performance and that their abilities to visually attend to task-relevant cues is more related to spatial performance.

One issue in the extant literature is that different methodologies have led to different conclusions regarding the role of language in the development of children's spatial skills. On the one hand, experimental studies testing the role of spatial language on performance show that providing children with spatial language cues is more effective than providing them with visual cues in improving their spatial performance (Dessalegn & Landau, 2008; Miller et al., 2016). These findings suggest that visual attention may not be central in supporting spatial performance. On the other hand, individual difference studies suggest that children's attention to task-relevant information underlies their spatial performance and that this attention is not specific to language (Miller & Simmering, 2018; Miller et al., 2017). Similarly, studies assessing children at ages before they can produce spatial words, or assessing verbal recoding during the task on a surprise trial, also suggest that children are not explicitly using language to support their spatial performance. The former studies are limited in that they use only outcome measures and do not directly assess the effectiveness of their visual cues manipulations in drawing children's attention. It is possible that the verbal and visual cue manipulations were not equivalent in drawing children's attention and thus performance in the visual condition was worse because children did not attend as well to the manipulations. Children may have attended better to verbal cues than visual cues and this underlying difference in attention is what was driving differences in children's performance. The latter studies are limited in that they assessed both language use and attention outside of children's participation within the spatial tasks or did not directly test effects of language versus visual attention.

A limitation across all these studies is that there is limited understanding of how children visually attend to spatial information in the moment of a task. Understanding children's visual attention and how attention changes across verbal and visual cues manipulations will provide novel insights into the processes by which language and/or visual attention support children's spatial processing in the moment of a task. There is thus a need to use more direct measures of visual attention to understand relations between language and non-verbal attention over development.

Visual Attention and Relation to Spatial and Problem Solving Tasks

Investigating eye-gaze patterns has led to understandings of adults' visual attention during spatial tasks and of how visual attention relates to spatial performance (e.g., Andersen, Dahmani, Konishi, & Bohbot, 2012; Just & Carpenter, 1976; Nazareth et al., 2017). Additionally, investigating eye-gaze patterns has led to insights into how directing visual attention can improve spatial and problem solving performance (e.g., Bailey et al., 2012; Grant & Spivey, 2003). Most research in these domains has been conducted with adults but a few studies have directly tested visual attention in children (e.g., Hoffman, Landau, & Pagani, 2003; Thibaut & French, 2016). Although these studies were not investigating relations with language use, their results are informative in illustrating how visual attention relates to spatial reasoning and how visual attention can be directed without language to facilitate performance.

Eye-gaze patterns during tasks.

In the spatial domain, eye-gaze patterns have classically been studied using mental rotation tasks (Just & Carpenter, 1976) and have more recently been applied to investigating strategy use during navigation (e.g., Andersen et al., 2012; Viaene, Vansteenkiste, Lenoir, De Wulf, & De Maeyer, 2016). It is important to note that while eye-gaze patterns have been predictive of performance and strategy use across different types of tasks, the specific eye-gaze patterns that predict performance depend on the type of task. For example, studies have used the classic Shepard & Meltzler (1971) mental rotation task where participants are presented with two 3-D figures and have to indicate whether the two figures if rotated would be the same object. Eye-gaze patterns in this task are characterized by frequent saccades between similar segments of the two figures, followed by longer fixations while the participants are transforming and comparing the figures, followed by frequent saccades at the end when participants are checking their decisions (Nazareth et al., 2017). In contrast, in navigation tasks where one does not need to make frequent comparisons among items, we see that the number of fixations and duration of fixations are related to performance (e.g., Andersen et al., 2012). Considering differences across tasks in eye-gaze fixations, I focused my review of visual attention during spatial tasks with adults on navigation research, investigating how adults use landmarks to support their navigation, as these tasks are most similar to the type of task that were used in this dissertation.

Investigating adults' navigation reveals that the number of fixations and the time fixating on particular landmarks is related to whether adults use the landmarks for navigation and report that the landmarks were used. For example, Andersen et al. (2012) used a virtual maze task to investigate navigation strategies of previously visited routes. They found different patterns of eye-gaze for people who reported a spatial learning strategy, using landmarks to orient, compared to people who reported a response learning strategy, using sequences of body movement. Specifically, individuals who reported using a spatial learning strategy tended to have a higher number of fixations onto the landmarks compared to individuals employing a response learning strategy. This was also reflective of sex differences in the task as men used more response learning strategies while women used more spatial learning strategies. Additionally, for both men and women, as the number of landmarks in the environment increased, participants spent more time fixating on the landmarks. Similar kinds of eye-tracking studies that had participants navigate inside buildings show that the proportion of time fixating on particular landmarks was related to individuals' use of such landmarks (Ohm, Mueller, Ludwig, & Bienk, 2014; Viaene et al., 2016). These studies show that measuring the types of landmarks adults fixate on and the duration of their fixations is a reliable way of understanding how adults use landmarks in spatial reasoning.

In research with children, there has been less research investigating visual attention during spatial tasks or related problem solving tasks. The few studies measuring children's eye-gaze have found differences among children and adults or between typically and atypically developing children in the saccadic patterns to and from the different test items and in the amount of time fixating at particular test items (Hoffman et al., 2003; Thibaut & French, 2016). For example, Thibaut & French (2016) used a conceptual analogies task (a non-spatial task) in which children and adults saw two pictures in an A:B relation (e.g., mitten and hand) and then had to choose a picture for D, for which C:D (e.g., shoe and foot) shared the same relation as A:B. Five- and eight-year old children exhibited different strategies than adults. Children had more saccades between C and the response options while adults had more saccades between A and B. This shows that adults spent more time attending to the relevant relations in the task to form the analogy. French and Thibaut (2014) also used children's data in this task to create a neural network model based on saccade patterns. Their model of performance based only on saccade patterns predicted children's accuracy in the task up to 90%. This showed that, without taking into consideration conceptual knowledge for the task, the specific ways individuals attended to the stimuli was strongly related to task performance.

Additionally, studies show that participants with William Syndrome show different patterns of eye-gaze to matched control groups, which may underlie some of the spatial deficits exhibited by this group. For example, this effect has been tested in a block construction task in which children see a sample picture of how blocks go together and then are asked to put the blocks together to match the picture. In this task, children with William Syndrome showed fewer fixations to the model for complex puzzles and checked their puzzle against the model less frequently when checking their solutions compared to a matched control group (Hoffman et al., 2003). Together, these studies with eye-tracking have provided some insights into how children are allocating their attention to different task stimuli and the strategies they use to perform the tasks. However, unlike the studies with adults (especially with navigation), these studies do not convey how children are processing the specific relations or specific features characteristic of the stimuli that are fundamental for performing the task (e.g., how they process the information within test items in the analogy task or within the puzzle pieces in the block construction task). Thus there is a need to understand what kind of spatial cues children attend to while performing spatial tasks.

Manipulating eye-gaze patterns to improve performance.

In addition to findings that eye-gaze patterns relate to performance on spatial and problem solving tasks, research has also shown that directing visual attention can improve task performance, when modeled based on eye-gaze patterns of successful participants. For example, Grant & Spivey (2003) tracked participants' eye-gaze while performing the classic Duncker Radiation problem, in which participants had to identify how to use a laser to cure a person with an inoperable tumor (solution is to fire multiple low-intensity lasers outside the healthy skin tissue surrounding the tumor). The researchers showed that successful problem solvers tended to look more to the outer region of the diagram (skin area) and less to the inner region (the tumor). In a second experiment, the researchers directed participants' eye-gaze to the outer region of the diagram by having the outer region subtly pulsate and showed that participants in the pulsation condition were twice as likely to solve the task as those in control conditions. Thomas & Lleras (2007) followed up on these results. In their task, after having participants view the diagram, they had participants complete a tracking task in which they had to detect digits that appeared on the diagram, implicitly directing participants' eye-gaze to different regions of the diagram. In one condition, the tracking task had participants embody the solution to the problem, making several saccades to the outer region of the diagram. Participants whose eye gaze embodied the solution to the task performed better on the task, despite reporting no knowledge that there was a relation between the tracking task and the radiation problem. Similarly, within the spatial domain Bailey et al. (2012) used a subtle gaze direction technique, which subtly manipulated viewers' gaze while they performed tasks, directing their attention to relevant parts of an image. The researchers showed that subtly directing eye-gaze improved participants' accuracy in remembering spatial locations in a complex scene.

Similar effects of directing attention have been found with infants' visual discrimination and with conceptual analogy tasks. Successful performance in visual discrimination tasks in infancy is often characterized by infants who have many short looks to different regions of the screen, tending to scan more regions than infants who exhibit longer looks. Jankowski, Rose, & Feldman (2001) attempted to make longer lookers have shorter looks. The researchers illuminated different quadrants of the screen to encourage longer lookers to have shorter looks and look at more regions of the screen. This visual manipulation resulted in longer looking infants performing similarly to shorter looking infants. Additionally, in a conceptual analogies task when forcing children to look at the A:B pattern by covering up C: and the answer choice options, children performed better in the analogy task because they spent more time fixating on the relevant A:B pattern (Glady, Thibaut, French, & Blaye, 2012). These results show that implicitly directing eye-gaze can lead to improved performance for both adults and children on spatial and problem solving tasks, without explicitly using language or without participants' awareness of such effects.

In summary, research investigating eye-gaze patterns highlights the importance of directing attention to task-relevant information while performing spatial and problem solving tasks. Directing attention is predictive of performance on these tasks and predictive of the types of strategies one employs (e.g., Andersen et al., 2012; Thibaut & French, 2016). Additionally, both modeling eye-gaze patterns and directing eye-gaze in particular ways can lead to improvements in performance (e.g., Bailey et al., 2012; Grant & Spivey, 2003). While we cannot rule out the possibility that directing attention increases verbal encoding, these findings demonstrate that visual attention is intrinsically related to performance and simply directing attention can have a causal effect on performance.

Chapter 2: Current Study

The current study used eye-tracking to increase understanding of how language and visual attention processes relate to children's spatial reasoning. As described above, spatial language is a useful tool that can be used to support spatial reasoning and that can direct children to encode task-relevant information while they perform spatial tasks (e.g., Landau et al., 2010; Loewenstein & Gentner, 2005). Some research suggests that spatial language is unique in influencing children's spatial performance as visually highlighting task-relevant information does not improve children's spatial performance to the same extent as language (Dessalegn & Landau, 2008; Miller et al., 2016). In contrast, other research suggests that changes in spatial cognition can be explained by non-verbal changes in children's attention (Miller & Simmering, 2018) and that there is little evidence that children or adults spontaneously use verbal encoding to support their performance (Feist & Gentner, 2007; Hitch et al., 1988; Nardini et al., 2006). Additionally, research investigating eye-gaze patterns shows that eye-gaze patterns are a powerful predictor of performance and that directing eye-gaze to task-relevant information leads to performance improvements (e.g., Grant & Spivey, 2003; Thibaut & French, 2016). The contrast in the effects found across these studies in the existing literature raised important questions regarding the role of language and visual attention in the development of spatial skills.

This dissertation used a similar paradigm to past research (Miller et al., 2016) to address questions regarding the role of language versus visual attention in spatial skill development. In the study, 4- and 5-year-old children's eye-gaze was tracked while they participated in two spatial recall tasks, a production task, and a comprehension task. This age group was selected because there is high variability in children's spatial language use (Kuperman et al., 2012) and spatial skills (Verdine et al., 2017), as well as strong effects of language on spatial performance (Dessalegn &

Landau, 2013; Loewenstein & Gentner, 2005). The spatial task used assessed children's performance in using an intrinsic reference frame during spatial recall. An intrinsic reference frame, as described above, is defined relative to nearby objects such as the arrangement of objects on a desk or kitchen counter (Levinson, 2003). Each participant completed two spatial recall tasks: baseline recall task and cue manipulation recall task. In the baseline recall task, children were not provided with any specific cues referring to the landmarks. In the cue manipulation recall task, children were assigned to conditions of verbal cues, visual cues, and control. In the verbal condition, children were provided with verbal cues describing relevant spatial relations. In the visual condition, children were provided with analogous visual cues that highlighted these important relations. In the control children, children received no specific cues, similar to the baseline task. The recall tasks were modified slightly from the paradigm used by Miller et al. (2016).

This particular spatial skill was used because it is a basic spatial skill that is important for perceiving, remembering, and communicating about object locations (Levinson, 2003; X. Li, Carlson, Mou, Williams, & Miller, 2011; Mou & McNamara, 2002). Reliable selection of an intrinsic reference frame in recall emerges relatively late in development, between 5 and 6 years of age (Jensen, Miller, & Simmering, 2015; Nardini et al., 2006). This is surprising considering that infants and toddlers show evidence of using other types of allocentric reference frames, such as those defined relative to a single object or cues in a room (Bremner, 1978; Newcombe, Huttenlocher, Drummey, & Wiley, 1998). Four- and five-year-old children are in a developmental transition period in using an intrinsic reference frame as they can use this reference frame on some trials (Negen & Nardini, 2015) and under some supportive task contexts (Jensen et al., 2015) but cannot use it consistently across trials or contexts. As mentioned above, previous research using this task (Miller et al., 2016) found the effects of verbal and visual cues on performance that this dissertation aimed to test (verbal cues > visual cues ≥ control condition). Finally, this task was chosen because it is engaging for young children, shows ample variability to assess individual differences (Miller et al., 2016), and reliably yields similar patterns of performance across minor changes to the task structure (Jensen et al., 2015; Miller et al., 2016; Nardini et al., 2006).

In-between the two recall tasks, children completed a production task following Miller et al. (2016) and a break task. The production task was not used in any current analysis from the dissertation but will be used to answer future questions. The break task occurred after the production task to allow children to sit down and rest. Following the second recall task, children completed a comprehension task (following Miller et al., 2016) in which the experimenter provided children with a verbal cue and then the child needed to find the hidden toy. This task was used to ensure that the children understood the cues used in the study, as effects of language are only present when children understand the verbal cues (Farran & O'Leary, 2016) and may be used for future analyses beyond this dissertation.

This study investigated <u>whether and how visual and verbal cues provided by an experimenter and children's visual attention co-contribute to influence spatial performance</u>. Specifically, I compared children's eye-gaze patterns and search performance during the recall tasks when children were provided with cues (verbal versus visual) relative to receiving no specific cues (control and within-subjects baseline conditions). Comparing the cue conditions to baseline provides more power to detect effects of cue condition as it reduces effects of within participant variability. Additionally, comparing cue conditions to the control condition avoids conflating effects of cue condition with practice during the task. These comparisons assessed whether the cues: a) influenced children's spatial recall accuracy; b) influenced children's eye-gaze; and c) interacted with children's eye-gaze to influence spatial performance. I predicted that the study would replicate the spatial recall performance patterns (verbal cues > visual cues ≥ control) found in past research (Dessalegn & Landau, 2008; Miller et al., 2016) and would be consistent with multiple extant theories. However, different theories make different predictions regarding whether and how the conditions should influence eye-gaze patterns and interact with eye-gaze patterns to influence recall accuracy. There are three potential patterns of performance that align with the different theories on the role of language versus visual attention in spatial development. By investigating visual attention in this way I began to address why we see divergent findings in the extant literature that simply directing visual attention can increase task performance (e.g., Grant & Spivey, 2003), versus other findings that only language significantly augments performance (Dessalegn & Landau, 2008; Miller et al., 2016). While I focus on these three patterns of performance as likely possibilities based on what is currently known, it is important to note that more patterns are possible and may point to alternative explanations.

Pattern 1 (strong verbal encoding hypothesis) would show that only language cues and not attentional cues influence children's spatial recall. Children's eye-gaze could reflect similar levels of visual attention during the recall task across conditions with verbal and visual cues, as both manipulations are equally effective in drawing children's attention. However, due specifically to the different form of the cues, only children in the verbal condition would show a benefit from the added cues on recall performance. This pattern of results would support the idea that language provides children with qualitatively different information than visual cues: verbal cues augment cognition in a way that visual cues cannot. This pattern is most consistent with research suggesting that, without language, one cannot process spatial information to the same extent (e.g., Shusterman & Spelke, 2005). This pattern of results would be unlikely to occur considering ample research showing that children and adults can perform spatial tasks without using language (e.g., Feist & Gentner, 2007; Learmonth et al., 2001; Nardini et al., 2006; Ratliff & Newcombe, 2008). However, I considered this pattern because analogous results have been found in other domains, with comparable patterns of visual attention across cues that differentially affect performance. For example, in a study investigating infants learning spatial pairings of auditory and visual cues, R. Wu & Kirkham (2010) found that 8-month old infants' patterns of visual attention were similar across visual and social cue conditions, but only the social cues enhanced infants' learning.

Pattern 2 (weak verbal encoding hypothesis) would show that both language and visual attention influence children's spatial performance but through different mechanisms. Children would show different levels of visual attention across the cue conditions, with most attention drawn by the visual cues. Recall performance should be better with verbal cues. Children's performance in the verbal condition would not depend much on their fixation patterns. However, children's performance in the visual condition would depend on how successfully they directed their eye-gaze to the task-relevant information (i.e., landmarks). According to this pattern, language can provide a different format for representing task-relevant information and can reduce attentional demands required to perform a task. When language is not explicitly utilized during a task, one may need to exert more attentional resources. I interpreted this pattern as supporting a weak version of the verbal encoding hypothesis (e.g., Landau et al., 2010) in which language is beneficial but not necessary, although this theory has not been specified to this extent.

Pattern 2 could explain why we see divergent findings in the extant literature. When children are provided with language, they can use it as a separate code to reduce attentional demands and perform better on the task. However, without language children overall need to exert more attention and young children may have a difficult time directing their attention effectively to perform difficult spatial tasks. This pattern of results in which visual attention has a different effect on performance depending on the cues provided comes from research testing toddlers' learning from video and in-person interactions. This research found that toddlers' performance in the in-person condition was relatively unaffected by their visual attention. However, performance in the video condition was affected by visual attention; children who attended more had better performance (Kirkorian et al., 2016).

Pattern 3 (attentional hypothesis) would show that visual attention is the primary mechanism supporting spatial recall performance and language only augments recall through its influence on attention. Children would show different levels of attention across the cue conditions, with language cues directing attention more strongly than visual cues. Visual attention would strongly predict spatial recall accuracy, irrespective of the cues provided. According to this pattern, the specific type of cue is not important, but instead how strongly the cue directs visual attention is the underlying factor supporting performance. This pattern of results would most strongly align with research arguing that language is not unique in facilitating spatial performance (e.g., Miller & Simmering, 2018). According to this hypothesis, language is a highly effective cue for directing attention (e.g., Spivey, Tyler, Eberhard, & Tanenhaus, 2001) but is not qualitatively different than visual cues. This pattern 3 could also explain why we see divergent findings in the extant literature. In most prior studies, language may have been quantitatively more effective in drawing children's attention to task-relevant information than visual cues. A close correspondence between eye-gaze and performance on spatial and problem solving tasks has been found in research with adults showing that directing visual attention in task-relevant ways can increase task performance, without explicit verbal cues (e.g., Bailey et al., 2012; Grant & Spivey, 2003).

Chapter 3: Method

Participants

94 children (M_{age} = 4.88 years, *range* = 4.00 to 5.87, 43 females) participated in the study. Of these participants, 16 contributed no eye-gaze data due to lack of eye-tracker recordings (3), inability to calibrate eye-gaze videos (6), or other issues related to eye-tracking quality (8, criteria described below in Eye-Gaze Data Quality Analysis and Exclusion section). Participants contributing no eye-gaze data were included in analyses of search performance but were excluded in the eye-gaze analyses. There were 29 participants in the control (M_{age} = 4.89 years, 13 females), 32 in the verbal cues (M_{age} = 4.86 years, 4.86, 14 females), and 33 in the visual cues (M_{age} = 4.94 years, 15 females) conditions. Of the participants contributing eye-gaze data, there were 24 in the control, 26 in the verbal, and 28 in the visual conditions.

An additional 15 children participated but were excluded from all analyses for being non-compliant (9); reaching criterion for being multilingual (3)[1]; and failing to reach performance criterion in the recall tasks (3, described below in Search Scoring and Exclusions section). Caregivers completed an optional demographics questionnaire (85%). From the data collected, participants were identified within the following racial/ethnic categories: White (80%), Hispanic (5%), Asian (4%), Black (3%), and other (6%). The majority of caregivers reported receiving a college degree or higher (68%). Participants were recruited from a database of families interested in research participation through Simmering's SPACE Lab and recruitment efforts by Kirkorian's Cognitive Development and Media lab. Families of participants received $10 compensation for participation and children received small prizes.

---

[1] Caregiver reported on the language questionnaire that child's caregivers spoke to them regularly in a language other than English and/or the child received more than 2 hours a day of exposure to a language other than English

Sample Size Rationale.

The sample size for this dissertation was selected because it is a sample size that has been used in past research using similar types of mixed-effect logistic regression models with interactions of continuous and categorical variables. Across 6 studies with children ranging from 2 to 5 years of age, the sample size, per group, had a mean of 26 and ranged from 18 to 40 children (Axelsson, Perry, Scott, & Horst, 2016; Bunger, Trueswell, & Papafragou, 2012; Chabal, Schroeder, & Marian, 2015; Haendler, Kliegl, & Adani, 2015; Järvikivi, Pyykkönen-Klauck, Schimke, Colonna, & Hemforth, 2014; Kirkorian et al., 2016).

Apparatus

The test apparatus included an array of cups and stuffed animal landmarks situated on a short rotating table (see Figure 1). Four blue cups served as hiding locations and were placed on top of the table in a square configuration. There were four unique stuffed animal landmarks (frog, cow, pig, bear) placed in between each cup. This layout allowed children to use the relative position of each cup and nearby landmarks, an intrinsic reference frame, to differentiate the hiding locations. One footprint on an 8.5 x 11 inch sheet of paper was placed on the floor of the testing room to indicate the position of the child to stand. A small toy served as the target for hiding events. A white cardboard board (22 x 20 inches) was used to block the child's view of the array during delay periods of the recall and comprehension tasks. The children were asked to hold this board in front of them (with an experimenter's help), which helped prevent children from touching the eye tracker during the delay period. A solid colored curtain hung from ceiling to floor in the surrounding task space to block external landmarks in the room such that the only cues in the room were the cues on the table. An overhead camera was mounted on the ceiling to record children's search responses. Figure 2 shows dimensions of the testing array and room. A demographics questionnaire was administered to caregivers to assess basic demographic information such as the child's race/ethnicity and parent's education (Appendix A1). The language background questionnaire was used to assess children's second language exposure (Appendix A2). Both questionnaires were administered on a laptop computer through Qualtrics software.

Figure 1. Layout of the testing array. Numbers indicate the specific hiding locations (not displayed in the task).

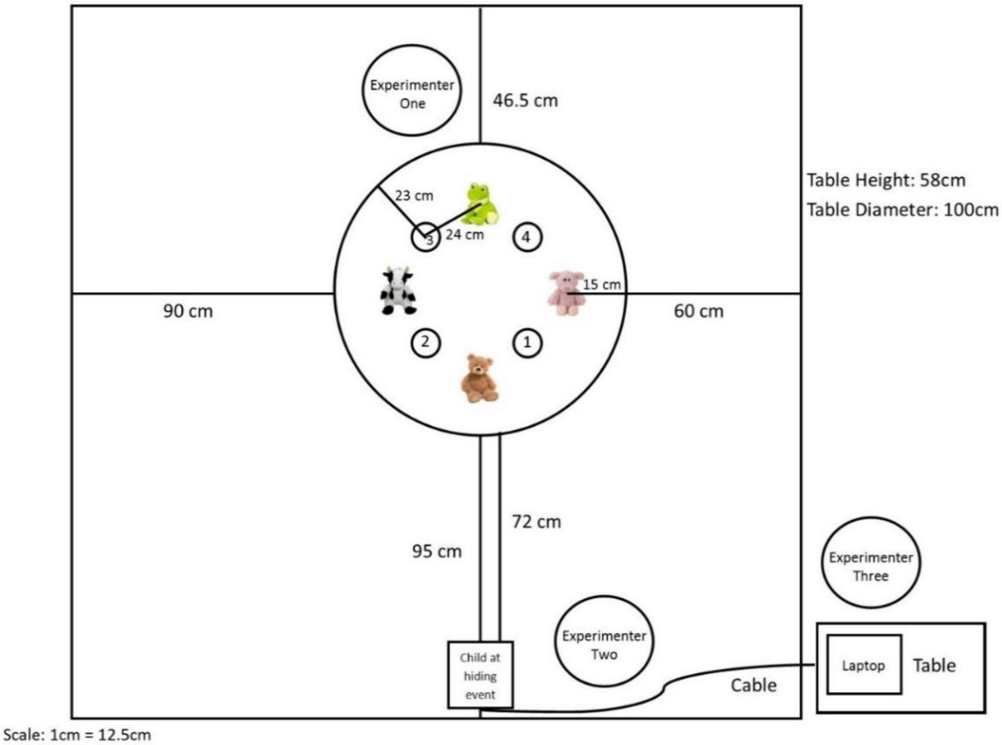


Figure 2. Layout and dimensions in testing room.

Children’s eye-gaze was tracked using the Positive Science child headgear model DB9-CHG. The eye-tracker was affixed to a stretchy cap and had a long tethered cord connected to a laptop computer. Children wore a small backpack that held the head gear power unit, which was also connected to the eye-tracker’s tether (Figure 3). The eye-tracker contained an eye camera, infrared emitting diode, and scene camera. The eye camera and infrared emitting diode were mounted on a bendable arm and pointed inward towards the child’s right eye. The infrared emitting diode illuminated the eye unobtrusively and created a corneal reflection to be used as a reference point. The scene camera was mounted on a headband in the middle of the child’s forehead and pointed outward. The scene camera had a standard field of view. The camera was the following dimensions (mm): width 3.60; height 2.70; diagonal 4.50; and focal length 3.50. The scene camera’s angle of view had degrees of: 54.43 width; 42.18 height; and 65.47 diagonal. Both cameras recorded at a rate of 30 frames per second. The eye-tracker can record with up to 0.5º accuracy, but accuracy typically ranges from 0.5º to 2º. Figure 4 shows a sample image from the eye and scene cameras.

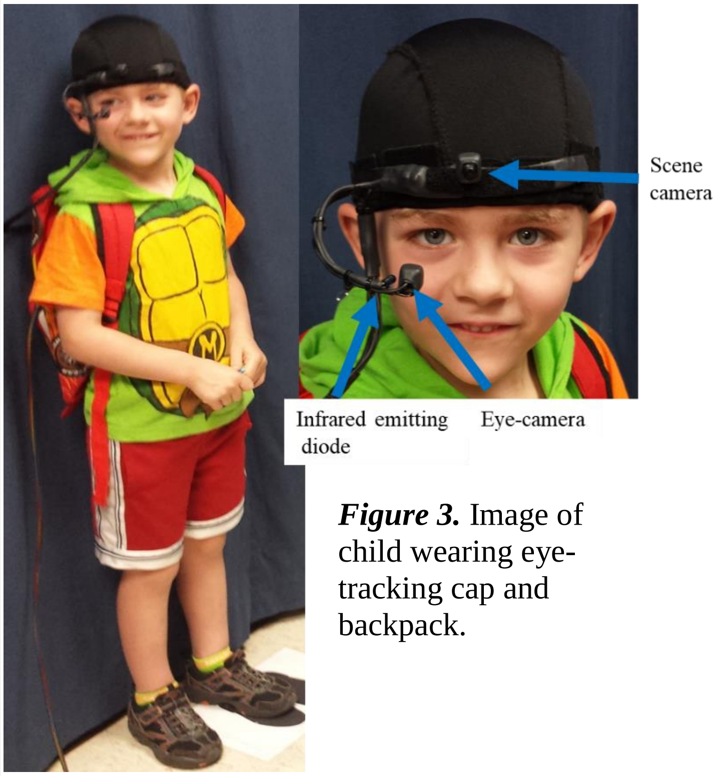


Figure 3. Image of child wearing eye-tracking cap and backpack.

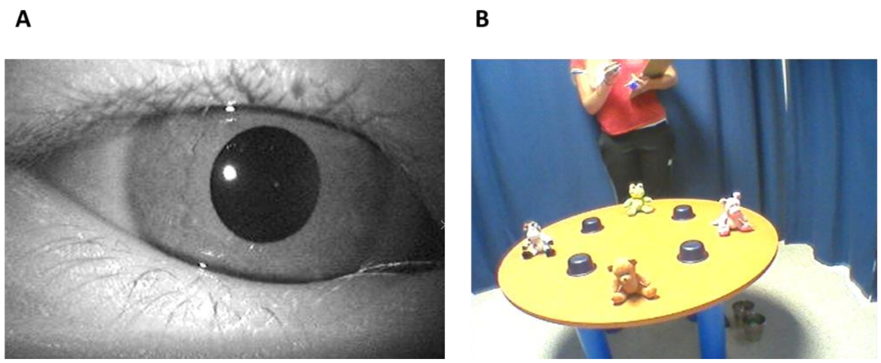


Figure 4. Sample image from the eye camera (A) and scene camera (B).

Design

The experiment involved three types of tasks using the rotating table array – recall, production, and comprehension – with an additional, unrelated task in the middle (see Table 1 for task objectives). Children completed two recall tasks (a baseline recall task and a cue manipulation recall task). The tasks were completed in the following order: baseline recall task, production task, break task, cue manipulation recall task, and comprehension task. The production task was administered between the two recall tasks to provide children with a break from the two tasks. Additionally, the production task was before the cue manipulation recall task so that the conditions did not influence children's production performance (cf. Miller et al., 2016). The comprehension task was administered last so that hearing the cues from the comprehension task did not influence children's recall task performance (Miller et al., 2016). The break task lasted 5 minutes and was administered on a tablet (unrelated to the research question) to give children time to sit down and rest halfway through the study. Note that I describe the procedures from the production task, but this task's analysis is beyond the scope of the dissertation.

Table 1

Task order and description.

Tasks (in order administered)	Objective	Number of trials	Calibration (before trials)
Baseline Recall	Recall with no cues provided	16 (4 no-rot., 12 rot.)	1, 5, 9, 13
Production	Production of hiding locations	8	1, 5
Break	Sit-down and rest	N/A	N/A
Cue Manipulation Recall	Recall under verbal, visual, control cues	16 (4 no-rot., 12 rot.)	1, 5, 9, 13
Comprehension	Comprehension of hiding descriptions	8	1, 5

Note. no-rot. = no rotation of table, rot. = rotation of table +90°, 180°, or -90°

Both recall tasks involved four rotation trial types, illustrated in Figure 5: no rotation (0º) or rotation (+ 90º, 180º, or -90º) of the array during the delay between hiding and search events. The no-rotation trials were used to ensure that children understood the task at the beginning of the study. They were also interleaved throughout the task to help the children stay motivated and to assess their motivation. In prior studies, children typically performed near ceiling on these no-rotation trials (Miller et al., 2016; Nardini et al., 2006). These trials were also not used in analyzing children's recall task performance. The rotation test trials were used to assess children's use of an intrinsic reference frame as, when the array rotated, the only visual features that remained consistent between hiding and search was the relative location of the target cup to other cups and landmarks on the array.

Children completed 16 trials per recall task (1 trial per hiding location and rotation type, with 12 rotation test trials and 4 no-rotation trials) divided into 4 blocks of 4 trials. Within each block, children completed one trial to each hiding location and each rotation type. The trials were randomized with the constraints that the same hiding location or rotation type were not used on consecutive trials. For the cue manipulation recall task, children were randomly assigned to one of three conditions: verbal, visual, or control. For the production and comprehension tasks, children completed 8 trials per task divided into 2 blocks. Within each block, children completed one trial to each hiding location. There were calibration phases at the start of each task and after every 4 trials (Table 1 for information on timing of calibration phases and see Calibration section below for details on calibration procedures).

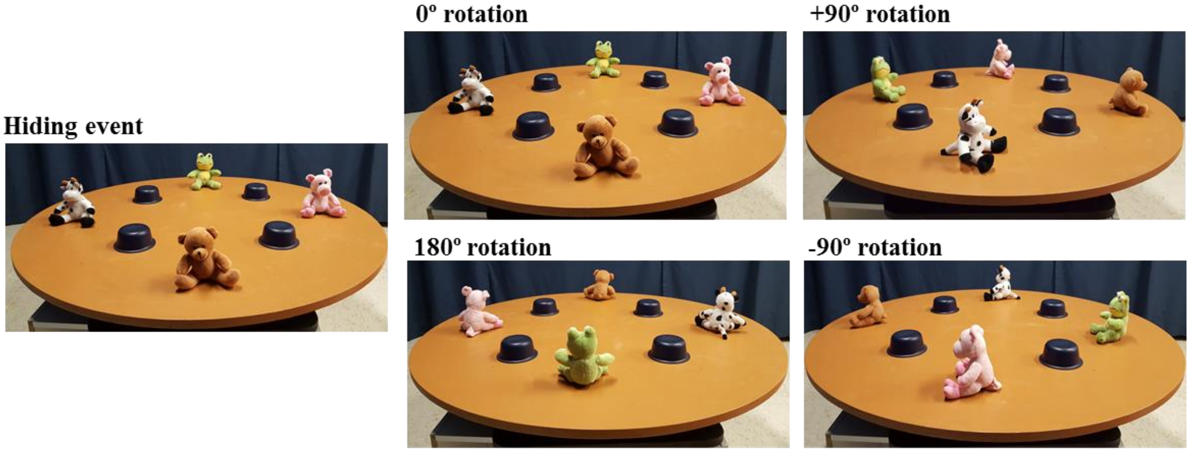


Figure 5. Diagram of testing array across rotation trial types. Displays children's view at hiding and at search events across 0º no rotation trial types, and +90º, 180º, -90º rotation test trial types.

There were three experimenters (E) present during all sessions. E1 provided children with instructions, ran the calibration phases, hid the toy, rotated the table, and marked children's responses during the study. E2 monitored the cord connected to the eye-tracker to ensure the child did not trip on the cord and helped children hold the white board during the delay periods. E3 (myself or a graduate student from the CDM Lab) put the eye-tracking cap on the child and monitored the computer connected to the eye-tracker to ensure that the eye and scene cameras were properly positioned. Figure 2 shows the location of each experimenter during the session.

Procedure

E3 obtained informed consent from the child’s caregiver and administered the demographics and language questionnaires (see Appendix A1 and A2). While in the waiting room, one of the experimenters read a short book to the child to familiarize them with the eye-tracking procedure. As shown in Appendix A3, the book showed pictures and described a child wearing the eye-tracking cap, referred to as the “Inspector Gadget hat,” and playing the “hiding and finding game” on the testing array. This book was intended to familiarize children with the eye-tracking cap and increase their comfort. The testing room was located in a separate room from the waiting room and children were asked whether they wanted their caregiver to accompany them during the study. Most caregivers stayed in the waiting room or in the testing room outside of the curtained space next to E3. A few caregivers accompanied the child inside the curtained spaced and stood to the left of the child in the corner, outside of the child’s primary viewpoint.

At the beginning of the session, E3 placed the eye-tracker cap on the child, positioned the headband with the scene camera between and just above the eyes, and adjusted the bendable arm, eye camera, and diode so that the eye was centered and illuminated. Throughout the session, E3 monitored the eye and scene camera on the laptop computer throughout the session to ensure that the eye camera had a full view of the eye and that there were no major changes to the positioning of the scene or eye camera during the session. The experimental task procedures are described first followed by the eye tracking calibration procedures.

Recall tasks

E1 started the baseline recall task by explaining the instructions to the child as follows: “In this game, I am going to hide this toy in one of these cups. After I hide the toy, you should remember where the toy is on the table, it will help you in this game. Then I will have you hold up this board and we will count to 5 together. When we say 5, you will put the board down and find the toy." E1 started the first trial and lifted the cup to hide the toy and simultaneously said "Look I am hiding the toy here." Then E1 told the child "remember look at the table" and then, approximately 2 seconds later, said "hold up the board and let's count to 5." This last cue and the 2 second delay before having the child hold up the board was intended to model the length of the verbal and visual cue conditions and provide children with more opportunity to scan the array. E1, E2, and the children then counted to 5 together (see Table 2). After the delay, E1 instructed the child to put down the board and find the toy ("Now put the board down and you can now look for the toy"). E2 held the board for the child during the entire session and helped the child raise and lower it. During search, E1 encouraged the child to find the toy on their first try but the child was allowed to search until they found the toy. On this first trial only (a no-rotation trial), the experimenter repeated the trial if the child chose the incorrect cup and repeated this trial until the child understood the task. The same procedure, with the exception of repeating the trial, was used on all subsequent no-rotation trial types.

For the first rotation test trial, the experimenter told the child before hiding the toy "now we will play the same game, but now when you hold up the board and we count to 5, I will turn the table and you will find the toy after the table turned." For subsequent trials before hiding the toy, the experimenter told the child whether the table did or did not move ("Now this time, I will/will not turn the table"), depending on whether the trial type was a no-rotation or rotation test trial (as in Miller et al., 2016). The same hiding procedures were used for the rotation test trials as were used for the no-rotation trials, except that the experimenter rotated the table +90º, 180º, or -90º during the delay and the experimenter reminded the child after the delay and before searching that the table moved ("Turn back around, you can now look for the toy. Remember I turned the table”). After the trial was complete, the experimenter moved the table back to its original position in the child’s view (Miller et al., 2016).

Table 2

Hiding event across conditions

Condition	**Instructions before task begins (only at start of task not on each trial)**	**Table rotation test trial instructions (each trial)**	**Hiding toy (each trial)**	**After hiding toy experimenter’s cues (each trial)**	**After hiding toy experimenter’s hands and gaze (each trial)**	**Start delay/child hold up board trial)**
Baseline and Control	“After I hide the toy, you should remember where the toy is on the table, it will help you in this game”	“Now this time, I will/will not turn the table”	“Look I am hiding the toy here”	“Remember look at the table” and then wait 2 seconds	Hands on clipboard, Looking at child	“Now hold up the board and let’s count to 5”
Verbal Cues	“After I hide the toy, you should listen to what I say, it will help you in this game”			“Between the bear and pig,” and then repeat cue	Hands on clipboard, looking at child	
Visual Cues	“After I hide the toy, you should look at what I am doing, it will help you in this game”			Lift and shake bear then lift and shake pig, and then repeat cue	Hands on landmarks, looking at the landmarks	

Note. The bear and pig are used as examples, but across hiding locations, the experimenter labeled or lifted and shook the two target landmarks (i.e., those closest to the hiding location; see Figure 1).

The cue manipulation recall task occurred after the production (described below) and break tasks. The same procedures were used for the cue manipulation recall task, except that children received different cues depending on their assigned condition (see Table 2). For the verbal condition, children heard the experimenter say a cue indicating the location of the toy relative to the two nearby landmarks and then repeated that cue one time. For example, if the toy was hidden under cup 1, the experimenter said “between the bear and pig, between the bear and pig”. The word “between” was chosen because it most precisely described the hiding location relative to the two landmarks. Loewenstein & Gentner (2005) showed that children benefit most when using precise spatial language. However, there is not much data on children’s production and understanding of the word “between.” This word is not included on the MCDI and thus is not in the word acquisition database Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2017). However, one study investigating children’s comprehension of the words “between” and “middle” found that 85% of 4- to 5-year-old children correctly drew an X mark in-between two objects after hearing the word “between” or “middle” (Hund, Bianchi, Winner, & Hesson-McInnis, 2017). In the same study 90% of parents indicated that their 4-year-old child understood the word “between”, and 95% indicated that their 5-year-old understood this word.

The visual condition paralleled the verbal condition. After the toy was hidden, the child saw the experimenter lift and shake the two landmarks nearby the hiding location one at a time and then repeated this manipulation. Note, the visual condition was different from that previously used (Miller et al., 2016). In the previous study, the experimenter moved the nearby landmark(s) on top of the target cup to represent the relation. However, this may have inadvertently disrupted children’s encoding of the spatial layout because the landmark position relative to the target cup changed during this manipulation. In the control condition, the experimenter ran the task the same way as was done in the baseline recall task with non-specific cues. The control condition was included in the experiment to account for any effects in the verbal and visual conditions that were due to learning during the study. Children’s cup selection during both recall tasks was recorded during the task by E1.

Production task.

The production task was administered between the two recall tasks. In our past research (Miller et al., 2016), we administered the production task before the recall task, and this order did not seem to affect children’s recall performance (i.e., similar performance to the recall task in another study that did not include the production task, Jensen et al., 2015). E1 explained the task to the child: “In this game, my friend [referring to E2] will turn around so that s/he cannot see the table. I will hide a toy in one of these cups and then we will have my friend turn back around. Then you will use your words and help my friend find the toy. In this game, you want to give my friend good clues so that s/he can find the toy.” On each trial, E2 turned his or her back to the table, then E1 hid the toy. After it was hidden, E2 turned back to the table and E1 prompted the child to describe where the toy was hidden. If the child pointed during the task, E1 instructed the child to use their words and hold onto the board (holding the board towards the ground). E1 prompted the child up to two times if the child did not give a verbal response or gave a vague response (e.g., “right there”, “I do not know”). After the child gave a response, E2 searched for the toy. E2 chose the cup they thought the child referred to; thus children received indirect feedback through search errors if their verbal response was unclear/ambiguous (as in Miller et al., 2016). An audio recorder was used to record the children’s verbal responses in this task (with the addition of an overhead camera and the eye-tracking camera as back-up recorders).

Comprehension task.

After the cue manipulation recall task, children completed the comprehension assessment as a manipulation check. The experimenter instructed the child to hold up the white board, blocking their view of the table, then E1 hid a toy in one of the cups. After hiding the toy, E1 instructed the child to put the board down and gave the child a verbal cue (e.g., “the toy is hidden between the bear and pig”). If the child did not find the toy on their first search, they were encouraged to keep searching. If participants in the verbal condition missed more than 3 trials (less than 62.5% correct), I planned to exclude their data from the analyses because they did not understand the cues used in the manipulation. However, no child in the verbal condition missed more than 3 trials. Although this criterion is only relevant for participants in the verbal condition, we had all participants run in this task to use for future analyses beyond this dissertation.

Calibration.

To enable subsequent offline eye-tracker calibration (described below in Offline Calibration), E1 drew children’s attention to various points on the testing array using a small bicycle light (Silicone Waterproof Super Frog LED Bicycle Head Front Light). The light was 3.2 cm wide and 1.0 cm tall and flashed a colored light on and off. Children chose what color light (white, yellow, red, blue, or green) they wanted to use. For the calibration phases, children were told that they would play a game called the “light game” and were told that the purpose of the game was to help the cap.

There were two calibration points for each cup and landmark on the array (16 total points). Each point was on the outer edge of its corresponding object, with one point next to the object on the table and the other point above the object (see Figure 6). For each object, the experimenter shined the light at the bottom point and then at the top point. E1 moved from point to point in an unpredictable order to keep the child engaged and to prevent the child from looking at the point before the light was flashing. The light was never shined on an object’s neighboring object in a consecutive order (e.g., cup 1 then Bear, cf. Figure 6). When shining the light at each point, E1 said, “Look here,” for the first point and then said, “Now look here,” for all the other points. E1 put the light behind their clipboard when not shining it at a calibration point. There were calibration phases at the start of each task and after every four trials throughout the entire study (for a total of 4 times during each recall task and two times during the production and comprehension tasks, see Table 1). Occasionally, there were added calibration phases if the eye-tracking cap moved during the study. In these cases, E3 adjusted the eye camera cap at the end of a given trial and then instructed E1 to conduct another calibration phase.

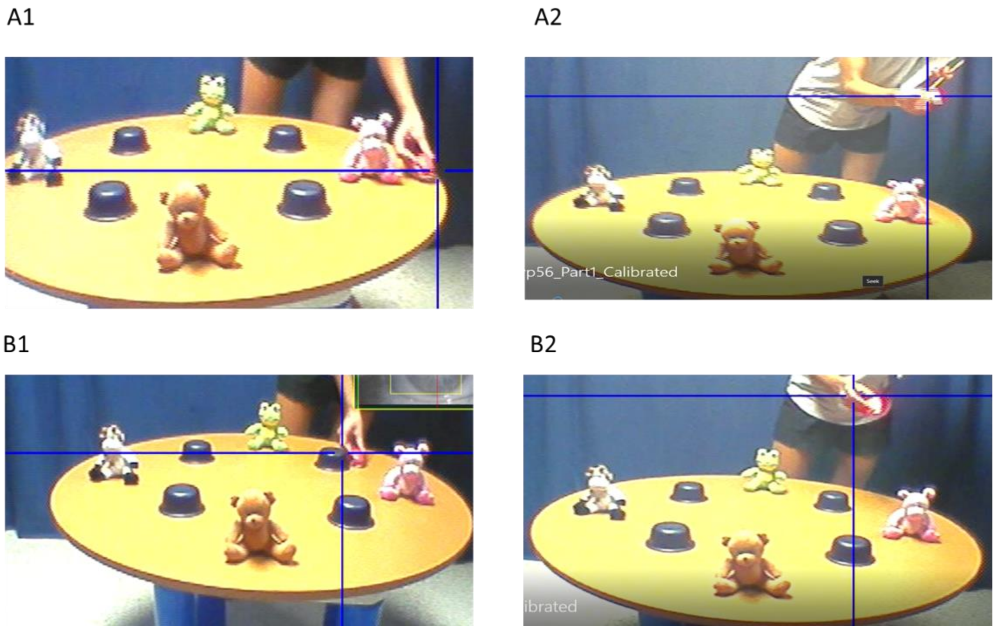


Figure 6. Sample calibration points (red bicycle light) in a rendered calibrated video with point of reference (blue crosshair). The top row represents calibration points next to a landmark for bottom points (A1) and top points (A2). The bottom row represents calibration points next to a cup for bottom points (B1) and top point (B2).

Method of Analysis

Here I will first describe the general analytic approach used in the statistical analyses and then describe recall accuracy data processing, followed by sections related to eye-gaze data processing. The majority of analyses test the effect of condition on a dependent variable (recall accuracy or landmark fixations) with subject as a random intercept. When recall accuracy was a dependent variable, I used mixed-effects logistic regression models in the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015) using the binomial family with the logit link function because the dependent variable was dichotomous (0 = incorrect, 1 = correct). When analyzing landmark fixations, a continuous variable, I used linear mixed-effects models, with Type III Wald F tests with Kenward-Roger degrees of freedom (cf. Kenward & Roger, 1997) using the lme4 package in R.

For comparisons across conditions, the condition variable was dummy coded treating baseline as the reference level. This method of analysis examines differences in performance in the control, verbal, and visual conditions relative to baseline performance. When evaluating change in performance from baseline to a respective condition, the analyses took into consideration participants as a random intercepts (i.e., each participant has its own change in intercept from baseline to its respective cue manipulation condition). Thus the analyses did not take into consideration whether baseline differed between groups (as there was no *a priori* reason to assume the groups differed, due to random assignment), but rather considered within-subject differences from each participant's performance during baseline to their respective cue manipulation condition.

Continuous variables in the models were mean centered, which included age and landmark fixations (when landmark fixations was not a dependent variable). Gender was contrast coded (-.5 female, .5 male). Block (1,2,3,4) was included in some models to test for effects of learning over time. Learning was assessed across blocks rather than individual trials because children's experiences were more similar across blocks than across trials (i.e., all children received the same three rotations test trials within each block but not in the same trial order). Block was coded with

3 orthogonal polynomial contrasts testing the linear, quadratic, and cubic effect (cf. Abelson & Prentice, 1997).

Search scoring and exclusions.

Children's cup selection in the recall and comprehension tasks were recorded during the task by E1 and by the overhead camera. Data from 23% (22/94 participants) of the recall and comprehension tasks were randomly selected to be coded offline for reliability. The offline coder was not an experimenter and was blind to E1's recorded responses. For the recall task, E1 hid the toy in the correct location on 99% of trials and reported the correct response on 100% of trials. For the comprehension task, E1 hid the toy in the correct hiding location on 100% of trials and reported the correct response on 99% of trials. Disagreements between E1 and the coders were resolved by the author and errors were corrected in the data.

Participants' performance on the no-rotation test trials was evaluated relative to chance as an indication that they were on-task. All no-rotation trials after the first practice trial of each recall task was evaluated for whether children performed above chance. Three participants performed below chance across both tasks (missing at least 5/6 trials) and were excluded among the 15 participants reported as excluded from all analyses (as listed in Participant section above).

Offline calibration.

Yarbus software (Positive Science, LLC) was used to synchronize, calibrate, and render the eye-gaze data after the session. All videos were calibrated by the author. Yarbus uses estimates of the center of the pupil and corneal reflection to map point of gaze onto the scene camera. To calibrate the video, the author placed calibration points when the bike light was flashing and the child clearly moved their eyes in the direction of the light. The calibration was conducted on the initial 16 points during the first calibration phase but was checked after every four trials at the new calibration points to ensure that the calibration was accurate. The accuracy was determined by a subjective rating by the author assessing whether the points of reference landed during the calibration phase on the bike light and during the hiding events on target cups when E1 hid the toy. The author watched the full video to spot-check the accuracy of the calibration. Since the calibration was conducted offline, the selected calibration points could be re-done or edited to improve the calibration. If the calibration was accurate for part of a video but inaccurate for another part due to factors such as the eye-tracker cap moving, the author set a new calibration in Yarbus for the appropriate video section. If the cap moved within a calibration phase before the next calibration phase, the trials were dropped due to inaccurate calibration (see Eye-Gaze Data Quality Analysis and Exclusion section below). Once all the calibration points were selected, the Yarbus software rendered the scene and eye video to create one calibrated video with a small cursor showing the point of reference on each frame in the scene video (cf. Figure 6) and an output file containing horizontal and vertical point-of-reference coordinates (relative to the scene video) and pupil diameter (height and width) for each frame.

Eye-gaze coding.

Dependent variables for the eye gaze data were determined in three phases. First, trained research assistants used Gazetag (Positive Science, LLC) to define visual fixations and code the location of each fixation during the hiding events. Second, a different set of research assistants used Datavyu video-annotation software to identify the approximate onset and offset frame for each hiding event. Third, the author used programs developed in Python 3.x for the purpose of this dissertation to refine the start and end times of each event, calculate dependent variables for each trial, and generate information about the quality of eye-tracking data. The eye-gaze coders were not one of the child's experimenters and were blind to the study hypotheses. However, the coders were aware of the experimental condition, as the manipulations made this knowledge unavoidable. Each session contained two videos: one video for the baseline recall task and one video for the cue manipulation recall task. Within each type of coding (fixation coding in Gazetag or trial duration coding in Datavyu) a single coder fully completed all steps for one of the two videos from a single participant's session. The same coder may have been assigned to code the other video from that participant's session or a different coder may have been assigned. For example, coder A completed all the fixation coding for one of participant Z's videos. For participant Z's other video, coder A or coder B may have been assigned to complete the fixation coding.

Fixation coding.

Gazetag was used for reducing the calibrated eye-gaze data into fixations and subsequently coding the fixations by labeling them. A fixation was defined in Gazetag with a minimum fixation time threshold of 100ms and a minimum frame distance of 10 pixels from the last fixation. There are two ways of coding fixations in Gazetag: linearly coding fixations and cluster coding, shown in Figure 7. When linearly coding fixations, the coder views both the scene and eye image and moves through the video in the order that the fixations occurred (Figure 7a). The video images are shown at the middle time point of the fixation and the coder has the option to play the full fixation as it occurred in the video. When using cluster coding, Gazetag creates fixation thumbnails at the middle of every fixation around the point of reference coordinate in the scene (Figure 7b and 7c). The thumbnails are grouped by similarity in color image statistics (Figure 7c).

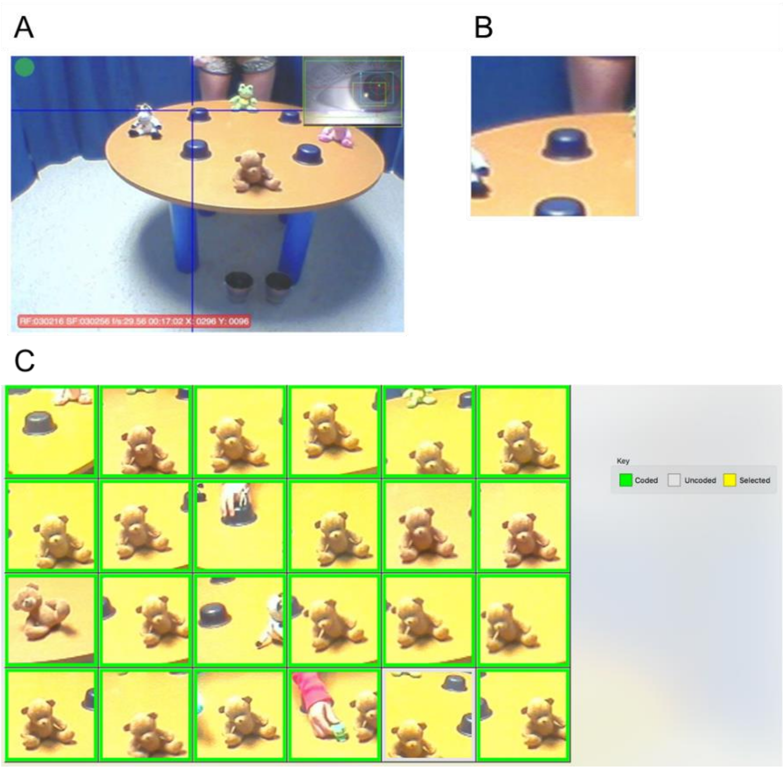


Figure 7. Top row shows images from Gazetag showing the same fixation using (A) the linear coding view and (B) the cluster coding view. In panel A, the blue cross-hair indicates the average point of reference for the fixation. In panel B, the center of the image reflects the average point of reference for the fixation. Bottom row shows one full panel of several thumbnails in cluster coding. Green box around thumbnails indicates which fixations are coded.

Coders labeled the fixations in four steps. A single coder completed all the steps for each video that they coded. In step 1, the coders linearly coded all fixations that did not occur during the hiding event (i.e., calibration phase, instructions, finding event). This reduced the number of fixations that needed to be coded for step 2. Additionally in step 1, the coders advanced one fixation at a time through the fixations during the hiding events and coded for instances of when the eye-tracker could not properly record the child’s gaze, decreasing the accuracy of the point of reference on the fixation. As seen in Figure 8, the eye image from the calibrated video contained a small green circle around the pupil and a small yellow cross around the corneal reflection from the infrared emitting diode. Coders were trained to look at the eye video for instances of unreliable eye-tracking, which occurred when the pupil tracking was off (i.e., green circle around the pupil was not centered inside the pupil), the corneal reflection was missing or the small yellow cross was not capturing the reflection light (cf. Figure 8B and 8C), or the child's eye was closed or in a squint. Coders were also trained to look for dramatic changes in the scene image, such as when the table suddenly looked slanted because the child's head was tilted. More detail regarding identifying instances of unreliable eye-tracking and how they influence the data are described below in the Eye-Gaze Data Quality Analysis and Exclusion section.

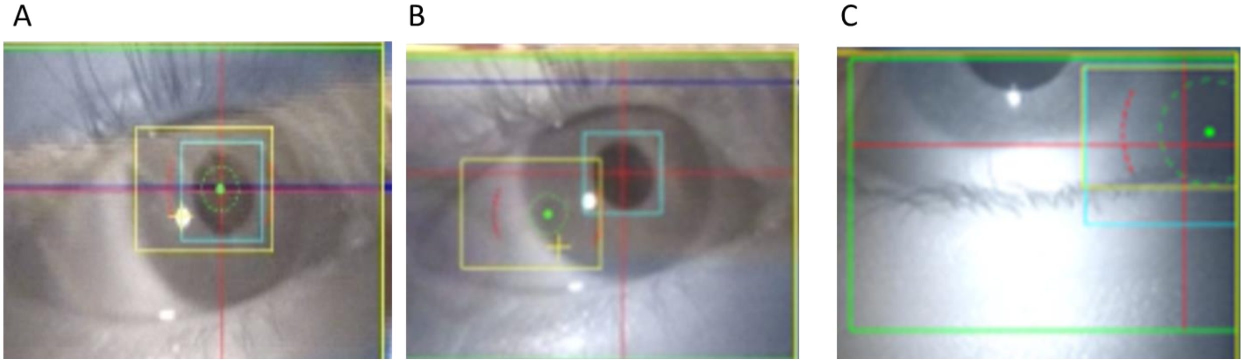


Figure 8. Enlarged images from the eye camera after calibration. Green circle tracks the pupil image and yellow cross tracks the corneal reflection from the infrared emitting diode. Panel A shows a well tracked eye-image. Panels B and C show poorly tracked eye-images, both images show lost pupil and corneal reflection tracking. Panel C occurred when the child was looking high upward. Notice that the size of the pupil track changed when it was inaccurately tracking the pupil.

In step 2, the coders used cluster coding to code the fixations occurring during the hiding events. The cluster coding setting indicated which fixations still needed to be coded or had already been coded during step 1. Research assistants coded the fixations that had not been coded in step 1 and labeled whether the fixations occurred on a particular cup, a particular landmark, the table (not specifically on a landmark or cup), the experimenter, or other locations in the room (i.e., floor, ceiling, and curtain). The coders coded the object that was at the center of the thumbnail (see Figure 7b and 7c). However, at times the fixated object was ambiguous from the thumbnail image, such as when there were two objects equally close to the center and it was unclear whether the fixation was on one of the objects or between the objects on the table. In the ambiguous cases, the coders left the thumbnail unlabeled (example shown in 7c, second thumbnail on the bottom right, thumbnail with no surrounding green box). In cluster coding, the research assistants were unaware of the specific time point within the trial when the fixation occurred.

In step 3, the coders coded any remaining unlabeled fixations in the linear coding mode, which was set to view only the unlabeled fixations. In this step, the coders had a sense of the time point in the video when the fixation occurred but were unaware of whether the child searched correctly on that trial. In the final step, step 4, coders viewed their coding library of coded fixation thumbnails, which showed the fixation thumbnails grouped by the label. In viewing the fixations, the coders identified when they mislabeled a fixation and identified when a labeled fixation looked different from most of the other fixations in the category. In these cases, the coders un-labeled the problematic fixations and then went back to step 3 to view those particular fixations and update the labels. Once coding was complete, Gazetag provided an output file showing the onset and offset time of each fixation, the duration of each fixation, and the point of reference (e.g., the name of the cup or landmark).

Trial duration coding.

A separate set of coders coded trial onset and offset times for each hiding event in Datavyu video-annotation software. Coders watched the video in real time to identify the approximate onset of a hiding event, then advanced the video frame-by-frame to determine the precise onset and offset of each trial of the hiding events. For trial onset, the coders identified the time point when E1 began to move their hand to hide the toy. Note that in my dissertation proposal, I proposed to begin the trials when the experimenter touched the target cup. I decided to change the coding in

Datavyu because I wanted to capture the child anticipating where the toy was going to be hidden, which could be anticipated from the direction E1's hand moved. Coding when the experimenter put their hand on the target cup sometimes would truncate when the child started fixating on the target cup prior to the experimenter touching it. However, coding when the hand moved towards the target cup was inexact because the experimenter made many movements and it was difficult to code the first movement downward towards the target cup. I thus decided to combine trial onset coding with fixation coding and used the child's first fixation to the target cup as the adjusted trial onset. A code was created in Python to adjust the trial onset time. Trials began when the child first identified the target cup's location after the experimenter started to move their hand downward to hide the toy (see Eye-Gaze Data Quality Analysis and Exclusion section below for dealing with trial onset issues such as when there were no fixations on the target cup). For trial offset, the coders identified the time point when the white board, which was used to block the children's view during the delay, was fully covering the child's view of the table. At times, the white board covered the array in the middle of a fixation; in these cases the fixation was shortened using Python so that the output file only contained the part of the fixation that took place before the trial ended according to trial duration coding. Once trial onset and offset times were identified, the Python code used a combination of fixation and trial duration coding to provide an output indicating the number and duration of each AOI occurring within the trial duration.

Eye-gaze measure rationale.

In the analyses, I considered using two measures of eye-gaze related to the proportion of time children spent fixating on target landmarks. For both measures, I chose to investigate fixations to target landmarks as the numerator (e.g., bear and pig for cup 1, see Figure 9) because children needed to use the location of the target landmarks on the array relative to the hiding locations to solve the task when the array rotated. I chose total on-task fixations as the denominator for these measures. All on-task fixations included everything the child fixated on during the trial, except for off-task fixations (ceiling, floor, and curtain), fixations on the white board, and fixations that were coded as unreliable eye-tracking in step 1 of fixation coding. There were a variety of different denominators that could have been chosen, such as total fixations on landmarks or total fixations on the cups and landmarks. I chose total on-task fixations because it is most informative of how much children fixated on the most relevant information versus other information they were processing during the task. For instance, if the denominator was out of all landmark fixations, two children may be similar in that their only landmark fixations were on the target landmarks but may be different in that one child spent more time and had more fixations on the target landmarks and the other child may have spent more time fixating on other information such as the experimenter. This measure would make the two children seem more similar even though there were differences in the extent to which they spent their on-task fixation times.

Figure 9. Layout of the testing array showing areas of interest to target landmarks for cup 1.

The first measure that I considered was the duration of fixations on the two target landmarks out of the total duration of all on-task fixations during the hiding events on each trial. The second measure was the same as the first measure except that it looked at the number of fixations to the target landmarks, instead of the proportion of time fixating on these landmarks. Due to limited research using these types of measures, it was unknown whether there would be differences in analyses using fixation number or durations. It was possible that the two measures capture similar variance in performance and the results may not differ based on the two measures. Children who spent more time looking at the target landmarks likely had more fixations to these landmarks. However, it is also possible that these measures capture different variance in performance. For example, children who had higher recall accuracy may have made fewer fixations but for longer durations on the target landmarks. Or children who had higher recall accuracy may have spent more time scanning the array and had many fixations on the target landmarks but for shorter durations.

I thus conducted two different analyses to investigate whether and how the two different eye-gaze measures influenced the results. First, I investigated which measure was most predictive of recall performance in the baseline condition prior to conducting the planned analyses described below. Specifically, I conducted two mixed-effect logistic regression analyses testing the probability of correct search on: a) proportion of time fixating the target landmarks during each hiding event/total duration of all on-task fixations during each hiding event; and b) number of fixations to the target landmarks during each hiding event/total number of on-task fixations during each hiding event treating participant as a random effect. When conducting the analyses, I found that both eye-gaze measures significantly predicted children's errorless search performance in the baseline condition. However, fixation number was a stronger predictor ($\chi^2(1) = 15.989$, $p <.001$, OR = 6.65) than fixation duration ($\chi^2(1) = 10.70$, $p =.001$, OR = 4.58).

Next, I used descriptive statistics to examine whether the pattern of differences in fixating on landmarks across conditions remained stable across the two different eye-gaze measures. The experimental conditions were aligned in the same ordinal position across the two measures and there was much similarity in the means and standard deviations across conditions when using fixation number [Baseline (*m* = .14, *sd* = .17), Control (*m* = .15, *sd* = .17), Verbal (*m* = .20, *sd* = .17), Visual (*m* = .58, *sd* = .16)] and when using fixation duration [Baseline (*m* = .12, *sd* = .17), Control (*m* = .13, *sd* = .17), Verbal (*m* = .18, *sd* =.17), Visual (*m* = .59, *sd* =.17)]. Based upon this data exploration, I used fixation number for my primary data analyses but conducted the same analysis with fixation duration and found the same pattern of results (Appendix C, Tables C2-C3, and Figures C2-C4).

Reliability for eye-gaze data.

A second coder coded 24% of participants' data (out of the 85 participants with calibrated video data) for fixation coding (25 baseline recall task, and 16 cue manipulation recall task) and for trial duration coding (18 baseline recall task and 23 cue manipulation recall task). Interrater reliability was evaluated using intraclass correlations (ICC) looking at agreement between first and second coders. This eye-gaze measure was calculated for both number and duration of fixations. For the numerator, interrater reliability was conducted on the proportion fixation to the target landmarks per trial, with fixation number having an ICC = .96, and duration having an ICC = .96.

The denominator was influenced by both fixation coding and trial duration coding. Fixation coding determined the first fixation on the target cup, unreliable eye-tracking fixations, white board fixations, and off-task fixations. Trial duration coding determined the range in which the first target cup fixation could occur (i.e., occured after E1's hand moved towards the target cup) and determined the trial offset. Thus for the denominator, I conducted interrater reliability twice, once for the duration determined by the fixation coders and once for the trial duration coders. The original coding was the same for both analyses (combination of fixation and trial duration coding), but the second coding differs based on how the fixation coders and trial duration coders specifically influenced these measures. For fixation coding of the denominator, reliability was ICC = .94 for fixation number and ICC = .95 for fixation duration. For trial duration coding of the denominator reliability was ICC = .99 (note that the reliability is the same for fixation number and duration since the coding is not based on fixations).

Eye-gaze data quality analysis and exclusion.

Analyses were conducted on the eye-gaze data to evaluate quality and identify outliers. Because head-mounted eye-tracking is a relatively new methodology, there are not yet well-established protocols for evaluating quality of such data. I first explain how eye-gaze data can be lost and/or how the point of reference on fixations is unreliable and then explain how I handled these issues within the data.

There are a variety of ways in which eye-gaze data can be lost or unreliable. First, the data can be lost from an inability to calibrate the eye-movements, resulting in no usable data from that participant's video or from specific trials within that video. Second, eye-gaze data can be lost from not being identified as a fixation. Fixations occur in between saccades (rapid movement of the eye in-between fixations); however, noise in the data input, which could be caused by loss of track of the eye-image, can prevent the software from properly identifying the fixation. Relatedly, fixations can be lost if the Yarbus software incorrectly measures the pupil. For instance, a loss of the corneal reflection as well as stark pupil movements to the corner of the eye (such as when the child is looking high upward or to the side) can result in lost or inaccurate measurement of the pupil (cf. Figure 8, differences in pupil tracking (green circle) across panels). Misidentifying the pupil not only results in fixations being lost, but can also result in fixations being captured with an inaccurate point of reference. Additionally, fixations might be lost due to difficulties in identifying smooth pursuits (Robinson, 1965). Smooth pursuits are smooth eye-movements that follow moving objects. The eye-tracker used in the current study identified fixations based on the x-y coordinates of the point of reference in the scene video rather than velocity of the eye-movement. Lastly, drastic changes in the child's posture from their posture during calibration, such as by the child tilting their head sideways or leaning far backwards, can result in the eye-tracker not properly capturing the point of reference.

As reported in the Participant Section, all data from six participants was dropped initially due to an inability to calibrate their eye-gaze video; either Yarbus failed to find the child's pupil on a large portion of the video (missing reflection light and/or too dark eye-image) or the child squinted often throughout the session. Of the participants' data that was successfully calibrated, 16% of the data was excluded due to indications of unreliable eye-tracking quality (this included 8 participants who were fully dropped due to having too many trials with unreliable eye-tracking quality issues, described below). A list of trial exclusions by condition can be found in Appendix B Table B1. First, individual trials were dropped when a segment of the data was identified as poorly calibrated (Appendix B Table B1 row 8, 2% of trials), due to the eye-tracker moving during the session before E3 could adjust the eye-tracker and E1 could conduct a new calibration phase. Trials were also excluded due to lack of fixations recorded on the target cup resulting in an inability to define trial onset (Appendix B Table 1 rows 9 and 10, 2% of trials). Some of these trials were dropped due to the child being distracted when the toy was hidden and others were dropped due to the child's point of reference onto the target cup not registering as a fixation in Gazetag.

After these initial exclusions, four analyses were conducted to evaluate eye-gaze quality and identify quality outliers. Note that these four additional analyses occurred after reliability analyses were conducted. See Appendix B Table B2 for trial quality statistics with eye-gaze quality outliers included and Table 3 (below) with them excluded. As seen in Appendix B Table B1, many trials were identified as outliers based on more than one of the following exclusionary analyses. First, trials were evaluated based on the duration of fixations captured out of the total trial duration (i.e., the proportion of all data points that were included in a fixation). Trials with unusually low fixation durations typically resulted from an inability of Gazetag to identify the fixations due to one or more of the unreliable eye-tracking metrics identified above. A visual inspection of the data, as seen in the histogram in Appendix B Figure B1A, shows that trials with less than 33% of the trial duration captured by fixations (more than 1.5 standard deviations below the mean) appeared deviant and were therefore excluded (Appendix B Table B1 rows 11, 15, 16, 17, and 18, 7% of trials[2]).

Second, trials were evaluated based on the pupil diameter size estimates from Yarbus from each frame captured within a trial (within or outside of a fixation defined by Gazetag). A deviant pupil size can result when Yarbus correctly identifies that a child is not fixating (such as by blinking or closing their eyes). However, as described above, a deviant pupil size can be an indication that the Yarbus software was unable to properly capture the child's pupil size, which had the potential of resulting in missed fixations or in fixations on which the point of reference was off (see Figure 8B and 8C for examples of misidentifying pupil). In this study, deviant pupil sizes often resulted in cases when the eyelid covered part of the child's eye (squinting and look downward towards the floor), when the child looked high upward towards the experimenter's face or towards the ceiling, and from Yarbus software otherwise losing track of the pupil. Pupil size was evaluated on each frame within a trial for whether the pupil diameter width and diameter

---

[2] Note, the trial drop percentages reported in the text are not mutually exclusive. Some trials were dropped for more than one reason and thus are included in more than one percentage. Table B1 found in Appendix B shows mutually exclusive percentages.

height was more than two standard deviations away from the average for that participant's trial. The pupil size was evaluated on the trial level rather than on the participant level because the eye-tracker's measures of pupil size could change when the eye-tracker moved slightly and slight movements are expected when the child moves to retrieve the toy or when E3 adjusted the child's eye-tracking cap.

Exclusions for the pupil diameter size measure were made on trials with more than 16% of frames indicated as having deviant pupil diameters (more than 2.5 standard deviations, Appendix B Figure B1B and Table B1 rows 12, 15, and 18, 3% of trials). This exclusion criterion was made on a slightly less conservative basis than the fixation duration captured criterion because a qualitative inspection of these outliers revealed that deviant pupil diameter readings for a short segment of the trial often occurred when the child was not fixating on one of the array's objects and was off-task, which does not as heavily influence the outcome of the results. The instances often occurred when the child was intentionally closing their eyes (distracting self), blinking, looking down at the floor. However, trials with more than 16% of the frames including poor pupil size often indicate issues that would more substantially influence the data, such as the eye-tracker losing track of the pupil (e.g., if the image was too dark). Additionally, data with a high percentage of deviant pupil diameters tend to have fewer fixations captured.

Third, trials were evaluated subjectively by fixation coders on the quality of the fixations captured. Trials were evaluated for the percentage of fixations that were coded in step 1 of the fixation coding as subjectively unreliable eye-tracking and all trials with more than 10% of the total fixations on the trial coded as unreliable eye-tracking (more than 2 standard deviations from the mean, Appendix B Figure B1C) were dropped (Appendix B Table B1 rows 13, 16, 18, 3% of trials). Lastly, since trial onset is related to the child's first fixation on the target cup, lag time between when E1 started moving their hand towards the target cup as coded by the trial onset coders and the child’s first fixation on the target cup as coded by fixation coders was evaluated. Trained research assistants examined why some trials had an excessively long lag (more than 2.5s) and dropped trials in which the eye-tracker failed to register the child’s first fixation on the target cup (Appendix B Table B1 rows 14 and 17, 1% of trials). Other instances of long lags, not resulting in trial exclusions, occurred when the child spent a long time looking at the experimenter or was a little distracted and looked away but still fixated on the target cup while E1 was in the process of hiding the toy.

Table 3

Eye-gaze quality statistics after quality exclusions

		Baseline			**Cue manipulation**			**All Trials**
Measure	**Unit**	**Control**	**Verbal**	**Visual**	**Control**	**Verbal**	**Visual**	**Total**
Fixation duration (ms)	*m*	3640.74	4398.46	4260.91	3769.86	4878.43	7640.56	4821.48
	sd	1197.90	1369.37	1230.24	1342.59	1478.76	2098.51	2036.11
Trial duration (ms)	*m*	6016.00	6651.73	6360.48	5947.28	8079.99	11954.4	7577.7
	sd	1659.33	1895.3	1537.79	1642.7	1986.03	2891.58	2953.31
Fixation duration/trial duration	*m*	0.61	0.67	0.67	0.64	0.61	0.64	0.64
	sd	0.15	0.15	0.13	0.15	0.15	0.13	0.15
Fixation number	*m*	10.91	11.99	11.35	10.58	14.35	22.73	13.81
	sd	4.5	4.83	4.45	4.25	5.48	6.97	6.78
Good pupil frames/ all frames	*m*	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	sd	0.03	0.03	0.02	0.03	0.03	0.02	0.03
Unreliable eye-tracking fixations/ al fixations	*m*	<.01	<.01	<.01	<.01	<.01	<.01	<.01
	sd	0.01	0.01	0.01	0.02	0.01	0.01	0.01

Note. Baseline is reported separately for each group, but analyses did not compare group differences within the baseline task (see description in Method of Analysis).

After the trials were identified as outliers and excluded, analyses were conducted on the number of rotation test trials included per participant. The mean number of rotation test trials included per participant was 10.41 (*sd* = 1.71) for the baseline recall task and 10.15 (*sd* = 1.44) for the cue manipulation recall task (out of 12 trials per task). Eight participants were excluded due to having 6 or fewer trials included in at least one of the recall tasks. As reported in the Results section and Appendix C Table C1 and Figure C1, the participant and trial exclusions due to eye-tracking loss did not result in significant change between samples in estimates of the effects of experimental condition on recall performance.

Additionally, an eye-tracking quality analysis was conducted, after the participants and trials were dropped, to examine whether there were differences across conditions in the proportion of the trial duration captured by fixations (i.e., fixation duration/trial duration). I conducted a mixed effects linear regression model with condition as a fixed effect, participant as a random effect, and proportion of trial captured by fixations as the dependent variable. As seen in row 3 of Table 3, there were only small differences across condition in the proportion of fixations captured per trial. However, the analysis found significant differences across conditions ($F_{3,\ 1547.78} = 15.50$, $p < .001$). Tukey post hoc tests revealed that all groups differed from each other except for baseline and control. There was more time captured by fixations in the baseline and control condition than in the verbal and visual conditions. Additionally, there was more time captured by fixations in the visual cues than verbal condition. These smaller proportion of fixations captured in the verbal condition could have resulted from the child looking upward at the experimenter (losing track of the pupil). In the visual condition this could have resulted in the child following the movements of the landmarks, creating a loss of data from an inability to capture smooth pursuit of eye-movements. To address the possibility that these difference may be affecting the pattern of results, I ran all analyses described in the Results section (below) that included eye-gaze measures while controlling for the proportion of time captured by fixations per trial (duration of fixations/trial duration) and found the same patterns of results. This data loss may be resulting in a slight underestimation of experimenter fixations in the verbal condition and in a slight underestimation of landmark fixations in the visual condition. However as shown in Figures 12 and 13 below, I found systematic and expected differences in children's fixation patterns across conditions despite these differences in proportion of fixations captured across conditions. These differences are important to consider when evaluating the results but these differences are likely not strongly influencing conclusions that can be drawn from the data.

A final quality analysis was conducted to assess whether there were significant differences in trial length. Although I attempted to make the trial length consistent across condition, trials in the visual condition tended to be longer due to the time it took E1 to move the landmarks. By contrast, in the baseline/control conditions, there was a short delay between the hiding event and the white board being lifted to block the child's view. Increasing the delay time without an obvious reason may have caused children to lose focus or become bored. In the verbal condition, the experimenters were told to speak slowly while giving the verbal cues, but it would be unnatural to extend the time to be comparable to the visual manipulation. An analysis with trial duration as the dependent variable, condition as a fixed effect and participant as a random effect indicated significant differences in trial length, $F(3, 1457.1) = 531.77$, $p < .001$. Tukey post-hoc test revealed significant differences across all conditions except baseline and control (i.e., visual > verbal > baseline = control). To evaluate whether these differences may have influenced the results of interest, I conducted all analyses described in the Results section (below) controlling for trial length and found the same pattern of results.

These two sets of analyses with controlling variables are important for testing whether some of the effects described in the Results section may have been influenced by characteristics of the data collection and processing that were not directly related to the central theoretical questions of interest. However, they are excluded from the remaining analyses because including them as predictors in the models creates issued with convergence and over-fitting. Dealing with convergence and over-fitting issues requires simplifying models. Since these factors do not heavily influence the results it made sense to exclude these factors.

Chapter 4: Results

The central goal of this dissertation was to increase understanding of the complex relations between visual attention and language in 4- and 5-year-old children's spatial skills. The analyses targeted the relation between eye-gaze and search performance, as well as the effects of the visual and verbal conditions. Three central questions were addressed: 1) the effects of condition on recall accuracy; 2) the effects of condition on eye-gaze; and 3) the effects of condition and eye-gaze on recall accuracy.

Effects of Cue Manipulations on Recall Accuracy

The first question addressed the effects of condition on children's recall accuracy. I tested whether the current study was consistent with previous research showing that spatial performance is enhanced most by experimenter provided verbal cues, with little or no benefit from visual cues (Dessalegn & Landau, 2008; Miller et al., 2016). I examined children's recall accuracy for the rotation test trials with all data (eye-tracking exclusions included). The model did not include interaction terms due to issues of convergence and overfitting the model.

Age and gender were included in the model to control for the effects of these variables. I predicted to find an effect of age as previous research has found transitions between 4- and 5-years of age in children's selection of an intrinsic reference frame (Jensen et al., 2015; Nardini et al., 2006). I did not have strong hypotheses regarding the effects of gender. There are sometimes effects of gender on spatial task performance, with boys outperforming girls (e.g., Levine, Huttenlocher, Taylor, & Langrock, 1999), but these effects have not been found in measurements of children's use of an intrinsic reference frame during recall (Miller et al., 2016; Nardini et al., 2006). I also included block to test for effects of learning over time. Of central interest, I predicted an effect of condition, with children performing best in the verbal condition and worst in the control and baseline conditions; the visual condition could be no different from control and baseline, or could fall in-between these and the verbal condition. These predictions are consistent with prior research and with all the hypotheses (strong verbal encoding, weak verbal encoding, and attentional) laid out in the Introduction.

In the model, the effect of age was significant, $\chi^2(1) = 14.65$, $p < .001$, OR = 1.95, 95% CI [1.38:2.77], but the effect of gender was non-significant, $\chi^2(1) = .04$, $p = .854$, OR = 1.03, 95% CI [0.75:1.41]. There was a significant overall effect of block, $\chi^2(3) = 49.32$, $p < .001$. Both the linear effect, $\chi^2(1) = 24.21$, $p < .001$, OR = 1.86, 95% CI [1.45:2.39], and the quadratic effects $\chi^2(1) = 29.09$, $p < .001$, OR = 1.66, 95% CI [1.38, 2.00] were significant but the cubic effect was not $\chi^2(1) = .76$, $p = .445$, OR = 1.10, 95% CI [0.86:1.41]. As seen in Figure 10 and found by the significant quadratic relation, children performed worse in blocks 1 and 4 and improved performance in blocks 2 and 3. This shows that children learned with practice and needed a few trials to reach their peak level of performance. However, children's performance tended to decrease during the final block, likely due to boredom or fatigue.

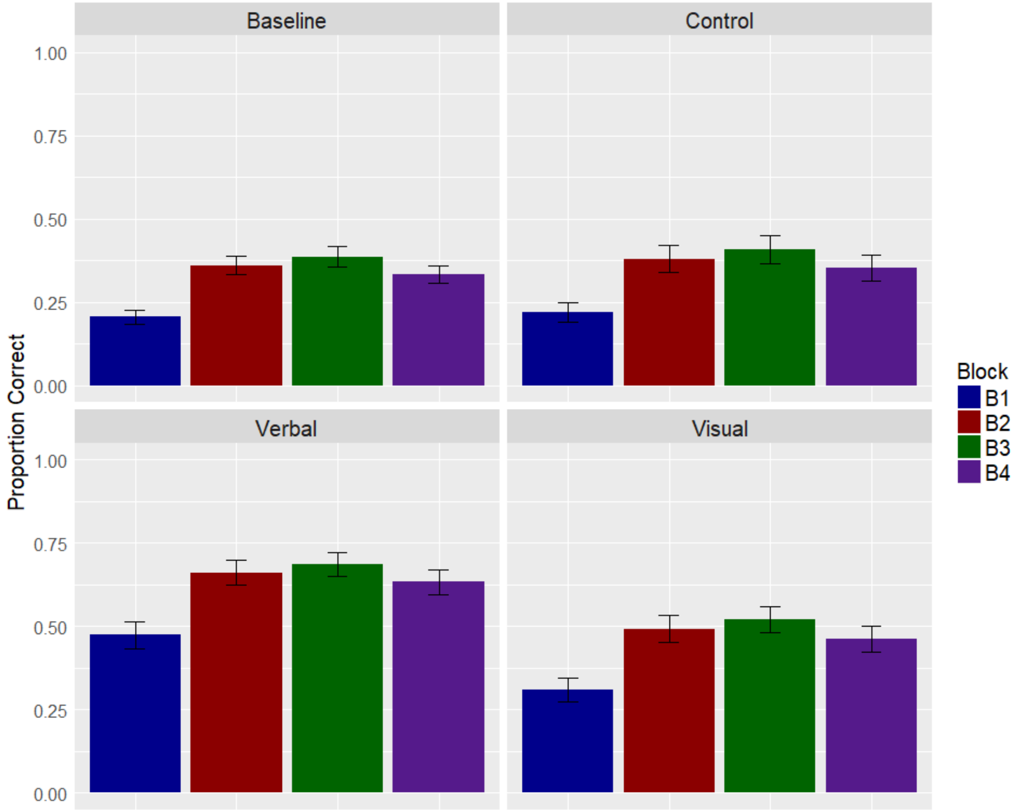


Figure 10. Proportion correct by condition and block, controlling for effects of age and gender. Error bars are plotted +/- 1 standard error for point estimates from the logistic mixed-effects model (black lines). Model plotted with all data, including data with eye-tracking exclusions.

Of central interest, the effect of condition was significant, $\chi^2(3) = 80.23$, $p < .001$, see Figure 10. Tukey postdoc tests were conducted to evaluate differences across conditions using the multcomp package in R (Hothorn, Bretz, & Westfall, 2008) appropriate for analyzing post hoc group differences in mixed-effect models. Relative to baseline performance, probability of correct search increased in the verbal and visual conditions (*ps* <.001), but not in the control condition ($p = .947$). Comparing across conditions, children had a higher probability of correct in the verbal condition than in the visual and control conditions (*ps* <.002). However, children in the visual condition did not have a significantly higher probability correct than those in the control condition ($p = .090$). These findings parallel the pattern of results found in Miller et al. (2016) using a different task layout and different number of trials, with children benefiting the most from language. Interestingly, children's performance improved relative to baseline in both the visual and verbal conditions, but the benefit to performance in the visual condition was not strong enough to differ significantly from the control condition.

All remaining analyses were conducted using the smaller sample (*n* = 77) of participants and trials following exclusions due to eye-tracking issues. As seen in Appendix C Table C1 and Figure C1, the pattern of recall accuracy across conditions remained the same with the smaller sample. Additionally, for the eye-gaze measure all following analyses were conducted using proportion of number of landmark fixations per trial but the same analyses are reported using proportion duration of landmark fixations per trial in Appendix C (Tables C2-C3, and Figure C2-C4).

Effects of Condition on Eye-Gaze

The second question examined how the conditions affected children's eye gaze. First, I examined the distribution of fixations across areas of interest for the various conditions. Second, I focused on whether the conditions significantly influenced children's eye-gaze patterns. As seen in Figure 11, based on descriptively examining means, children displayed different patterns of eye-gaze across conditions. In the baseline and control conditions, children tended to spend the majority of the trial fixating on the target cup and relatively little time fixating on the landmarks. Children spent the same proportion of time fixating on the different types of objects across the baseline and control conditions. In the verbal condition, children increased their fixations on the experimenter as well as on the target landmarks. This is likely due to the children paying attention to E1 providing verbal cues, and then referencing the landmarks associated with the cues. In the visual condition, children spent much less time fixating on the target cups and much more time fixating on the target landmarks. This suggests that children followed the visual cues of E1 shaking the landmarks during the trials.

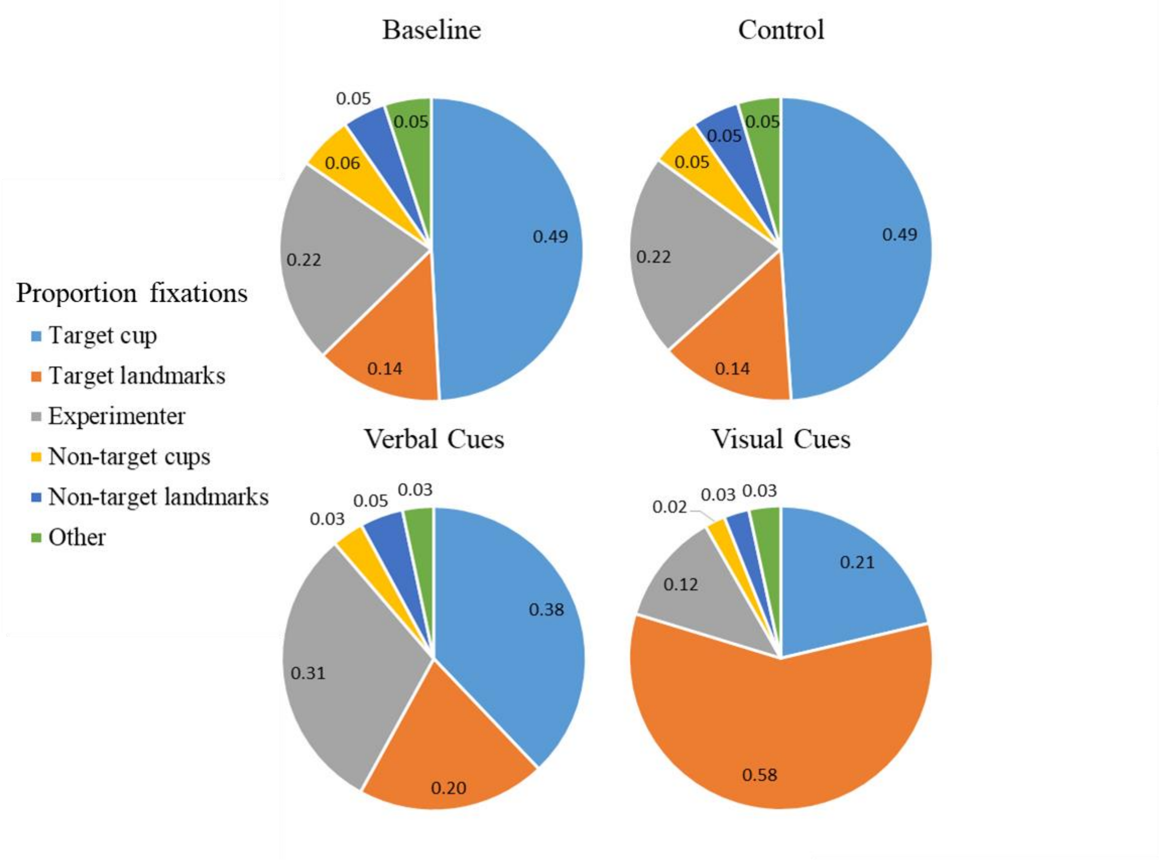


Figure 11. The relative proportion of the number of all on-task fixations on different objects in the testing room during hiding events across conditions.

Second, I tested whether the conditions significantly influenced the children's fixation on the target landmarks. I ran a linear mixed-effects model, using Type III Wald F tests with Kenward-Roger degrees of freedom (cf. Kenward & Roger, 1997) using the lme4 package in R. The model included proportion of target landmark fixations (number of fixations to target landmarks/ total on-task fixations) per trial as the dependent variable, with age, gender, block, and condition as fixed effects, and participant as a random effect. This analysis can begin to differentiate among the various hypotheses, but it is important to note that different patterns of eye-gaze do not exclusively support one hypothesis over the others. The strong verbal encoding hypothesis suggests no main effect of condition on children's eye-gaze, especially between the baseline/control and verbal conditions, as the verbal code provides an alternative to visualizing the relevant objects. The weak verbal encoding hypothesis suggests an effect of condition such that children should fixate more in the visual condition than in the verbal, control, and baseline conditions. The attentional hypothesis suggests an effect of condition as well: children should fixate more in the verbal condition than in the visual condition, with the visual condition being the same or slightly better than the baseline and control conditions. It is also possible that children will look more to the landmarks in the verbal and visual conditions than in the control and baseline conditions, because any form of highlighting the landmarks should increase fixations to the landmarks. This latter possibility could be consistent with the strong and weak verbal encoding hypotheses as long as eye-gaze is not influencing the effect of the verbal condition on recall performance. This would be inconsistent with the attentional hypothesis since that hypothesis posits that there will be more fixations on the target landmarks in the verbal condition, due to the verbal cues better directing children's attention.

In the model, the effect of age was significant, $b = .04$, $SE = .02$, $F(1,73.70) = 5.80$, $p = .02$, 95% CI [0.01:0.08]; while the effects of gender, $b = -.01$, $SE = .02$, $F_{1,73.22} = 0.35$, $p = .558$, 95% CI [-0.04:0.02] and block, $F_{3,1507.24} = 1.99$, $p = .113$, 95% CI [-0.02:0.03] were nonsignificant. Of interest, the effect of condition was significant, $F_{3,1365.24} = 455.65$, $p < .001$, see Figure 12. Tukey HSD post hoc tests demonstrated that the proportion of target landmark fixations per trial increased in both the verbal and visual conditions relative to baseline ($ps < .001$), but there was no difference for children in the control condition ($p = .99$). Additionally, children fixated most in the visual condition, followed by verbal, then control, with all pair-wise comparisons reaching significance (*ps* < .001). These results do not follow the patterns derived from strong verbal encoding, weak verbal encoding, or attentional hypotheses. These results suggest that the visual cues were effective in directing children's visual attention. Verbal cues affected children's target landmark fixations per trial, but not as strongly as the visual cues. These results are likely more consistent with a strong or weak verbal encoding hypothesis, but how all three hypotheses differ relates to relations with eye-gaze. Thus, these pattern of results raise the next question I addressed with my analyses, whether these changes in fixation patterns influenced children's recall performance.

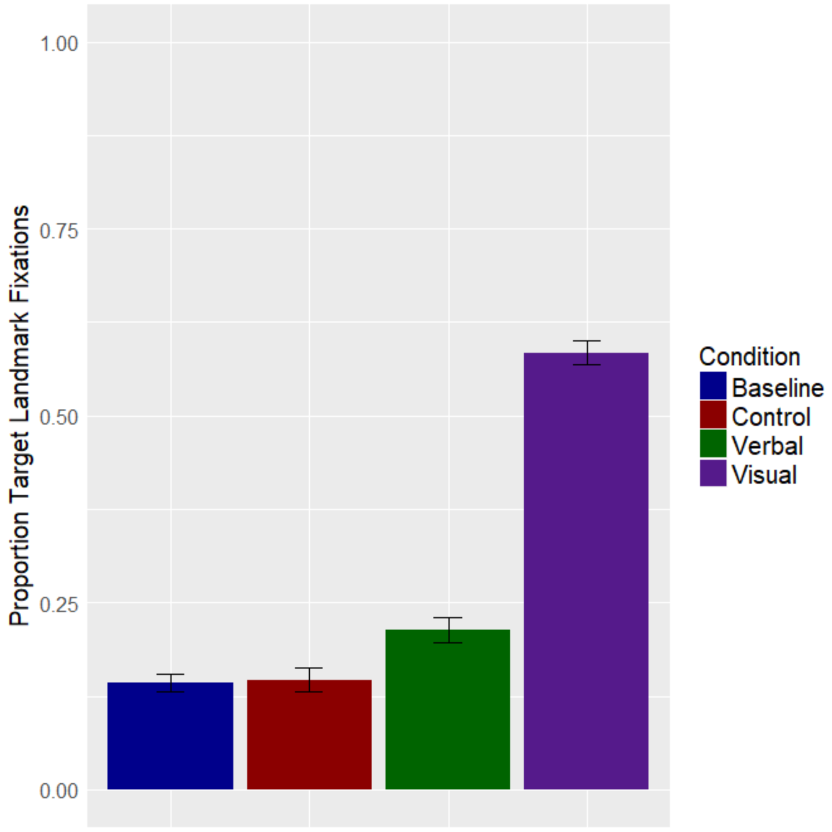


Figure 12. Proportion of target landmark fixations per trial by condition controlling for effects of age and gender. Proportion of target landmark fixations per trial was calculated as the number of fixation to target landmarks divided by the number of on-task fixations. Error bars show +/- 1 standard error for point estimates from the mixed-effects model.

Effects of Eye-Gaze and Condition on Recall Performance

Third and most central to the goal of understanding the complex relations between language and visual attention on children's intrinsic reference frame selection, I evaluated whether proportion of target landmark fixations per trial and the condition interacted to predict search performance. As described in the Eye-Gaze Measure Rationale section in Methods, the proportion of target landmark fixations per trial predicted children's recall performance in the baseline task, but it is unknown whether and how condition and proportion of landmark fixations per trial relate to predict probability correct. I used a mixed-effects logistic regression model to test the probability of correct search, with age, gender, condition, and proportion of target landmark fixations per trial as fixed effects and participant as a random effect. This model included the interaction between landmark fixation and condition. I removed block from this model because the model failed to converge with this variable included, and this measure was not related to my central questions. Additionally as reported in the analysis above, there was no effect of block on proportion of target landmark fixations per trial. I hypothesized three potential outcomes of this analysis, corresponding to the three theories described in the Introduction. The first possible outcome is that there was only an effect of condition, with the verbal condition being most predictive of performance, and no effect of eye-gaze or interaction (Figure 13A). This outcome would support a strong verbal encoding hypothesis showing that only language and not visual attention predicts performance. While included as a possibility, this finding is inconsistent with the findings that the proportion of children's target landmark fixations per trial in the baseline condition predict recall accuracy as well as inconsistent with past research (e.g., Bailey et al., 2012; Miller & Simmering, 2018).

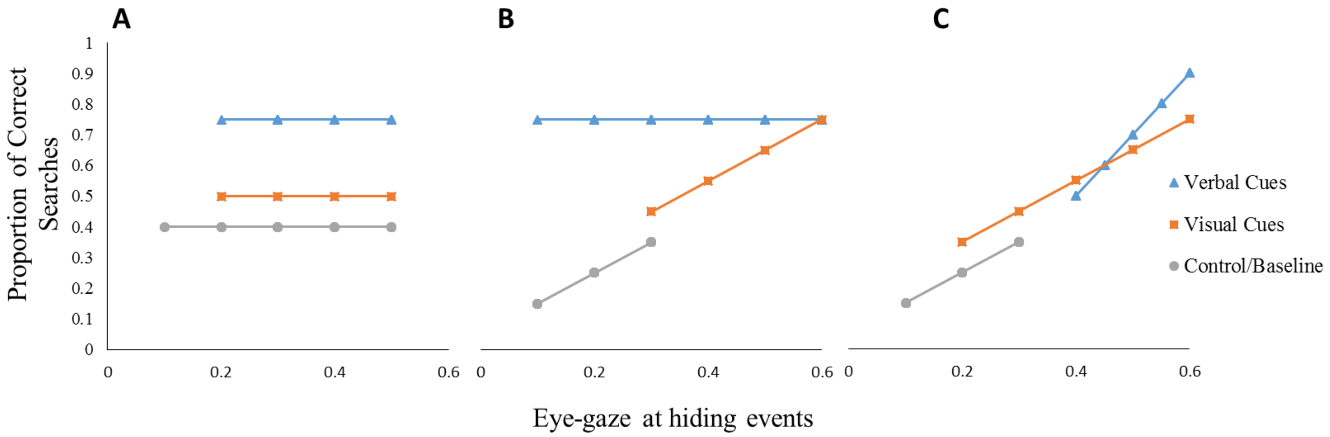


Figure 13. Hypothesized patterns of results under (A) strong verbal encoding hypothesis, (B) weak verbal encoding hypothesis, or (C) attentional hypothesis

The second possible outcome is that there is an effect of condition (verbal is best), no effect of eye-gaze (or a weak effect of eye-gaze), and an interaction between condition and eye-gaze (Figure 13B). For this outcome, there would be no relation between eye-gaze and search in the verbal condition: children could perform well no matter how much they fixated on the landmarks. However for the visual, control, and baseline conditions, children's looking behavior would predict search performance, with higher proportion of looking to the landmarks corresponding to more correct searches. This outcome would support the weak verbal encoding hypothesis, suggesting that language provides an additional code that children could use to perform the task without visually attending as much to the stimuli on the array. However, when not provided with verbal cues, children need to rely on visual attention to successfully solve the task.

The third possible outcome is that there is an effect of condition (verbal most predictive), and an effect of eye-gaze (more looking to the landmarks leading to better performance) but no interaction between eye-gaze and condition (Figure 13C). This outcome would suggest that children are attending more to the landmarks in the verbal condition, with children's overall attention to landmarks predicting performance. This third outcome would support the viewpoint that visual attention is underlying spatial performance. But considering that language cues were more effective in facilitating performance (here and in Miller et al., 2016), future work would be needed to evaluate whether any visual cues on their own could be as effective as verbal cues in supporting spatial performance. Based on my past research (Miller & Simmering, 2018), I predicted in my dissertation proposal to find this outcome that verbal cues are only more effective than visual cues due to their effect on attention.

In the model, there were significant effects of age, $\chi^2(1) = 12.80$, $p < .001$, OR = 1.86. 95% CI [1.32:2.64], condition $\chi^2(3) = 15.44$, $p = .002$, and proportion of target landmark fixations per trial, $\chi^2(1) = 11.44$, $p < .001$, OR = 5.12. 95% CI [2.00: 13.33]. As in the other models, there was no significant effect of gender, $\chi^2(1) = <.01$, $p = .962$ OR = .99. 95% CI [0.72:1.36]. Additionally, there was no interaction between condition and proportion of fixation on the target landmark, $\chi^2(3) = 4.48$, $p = .213$, see Figure 14. Tukey HSD post hoc tests following up on the effect of condition showed that children had a higher probability of getting a trial correct in the verbal condition relative to the baseline, control, and visual conditions (*ps* < .028). No other pair-wise comparisons reached significance (*ps* > .545). This is similar to the findings in the previous model testing the effects of condition on recall accuracy, when not controlling for proportion of target landmark fixations per trial, except that the difference between the visual and baseline conditions was significant in the previous model.

As seen in Figure 14, when controlling for effects of condition, the more that children fixated on the landmarks, the higher their probability of searching correctly. However, there was also unique variance associated with the verbal cue manipulation, beyond the effect of eye-gaze. When controlling for eye-gaze, children in the verbal condition had a higher probability of getting a trial correct than in all the other conditions. This suggests that language provided by an experimenter is not completely accounted for by children’s target landmark fixations, since the verbal cues condition uniquely predicted variance in recall performance. In the visual condition, children’s search performance increased relative to baseline; however, this effect disappeared when controlling for children’s target landmark fixations. This suggested that variance associated with probability correct in the visual condition was primarily captured by the effects of target landmark fixations.

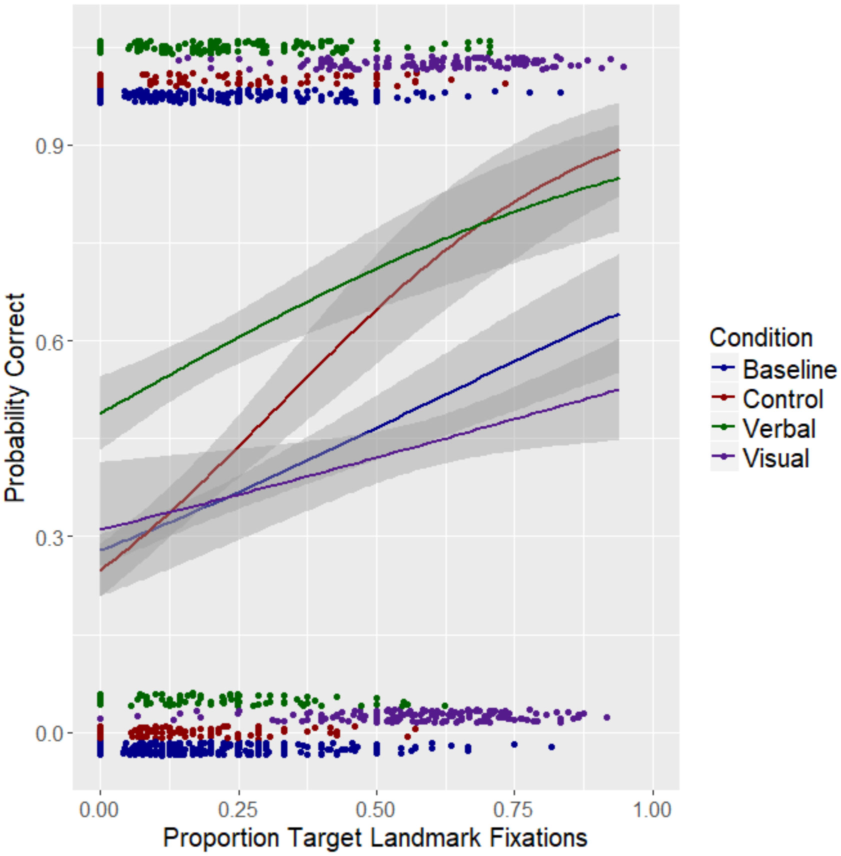


Figure 14. Probability correct by proportion of target landmark fixations per trial and condition, controlling for effects of age and gender. The data points represent individual trials for each condition indicating whether the trial was correct (1) or incorrect (0), with conditions offset vertically for visibility. Grey error bands are plotted +/- 1 standard error for point estimates from the logistic mixed-effects model.

To take these results one step further and directly understand whether some of the effects of the verbal and visual condition on probability correct resulted from changes in eye-gaze, I conducted mediation analyses. The three main statistical analyses conducted above fit the prerequisite standards for mediation (Baron & Kenny, 1986). Specifically for the verbal and visual conditions relative to baseline, a) condition had a significant effect on probability correct, b) condition had a significant effect on eye-gaze, and c) the effect of condition on probability correct was reduced when eye-gaze was added to the model. To conduct the mediation analyses, I used a bootstrapping procedure (Preacher & Hayes, 2004) in the mediation package in R (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014) using 10,000 simulations. For the verbal condition (Figure 15a), there was a small but significant indirect effects via proportion of target landmark fixations per trial $b = 0.03$, 95% CI [0.01:0.04], such that condition influenced eye-gaze to predict probability correct. As expected the direct effect, which is not accounted for by the mediation effect, was significant $b = 0.22$, 95% CI [0.11:0.32]. For the visual condition (Figure 15b), there was a significant indirect effect via eye-gaze, $b = 0.13$, 95% CI [.04:0.21] such that the visual condition influenced children's eye-gaze to predict recall accuracy. As expected, the direct effect was nonsignificant, $b = -0.05$, 95% CI [-0.17:0.06], as the effects of the visual condition on the probability correct was primarily due to the effects of fixations. These mediation analyses suggest that the part of the verbal condition effects on recall accuracy resulted from effects of the condition on children's eye gaze (as shown by the indirect effect); however, most of the effect arose from other factors related to verbal cues not attributed to eye-gaze (as shown by the beta estimates of the direct effect being larger than the indirect effect). For the visual condition, the mediation analyses suggest that the effect of the visual cues in supporting recall performance arose through directing children's eye-gaze on the target landmark fixations. Only the indirect effect was significant and not the direct effect.

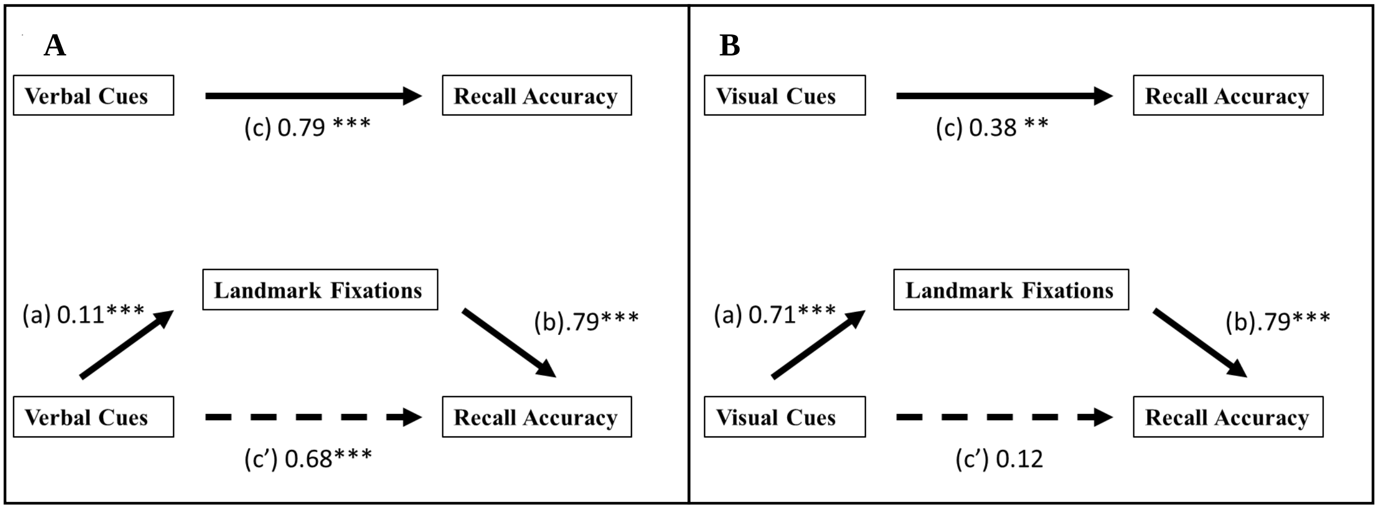


Figure 15. Standardized regression coefficients for the relation between verbal cues (A) visual cues (B) with baseline as the reference level and recall accuracy, mediated by proportion target landmark fixations, controlling for effects of age and gender.

Overall, these results are most consistent with a weak verbal encoding hypothesis. The results did not explicitly follow any of the patterns laid out in my predictions (cf. Figure 13). However, these results show that language uniquely supported spatial performance that was not directly tied to visual attention, suggesting that language can provide a unique code for representing spatial information. Importantly, these results also highlighted that visual attention with or without direct language support can enhance children's recall performance, demonstrating that there can be different routes to performing spatial tasks.

Chapter 5: Discussion

This dissertation was targeted towards understanding the complex relations between visual attention and language as these factors relate to children's spatial skill development. Past experimental studies demonstrated that language cues were more effective than visual cues in supporting children’s spatial performance (Dessalegn & Landau, 2008; Miller et al., 2016). However, individual difference studies found that young children’s attention to task-relevant cues is a stronger predictor of their spatial performance than is children’s use of task-relevant language (Miller & Simmering, 2018). Theorists have proposed various hypotheses to explain such effects of language versus visual attention on children’s spatial performance. The strong verbal encoding hypothesis suggests that language causes changes in children’s spatial cognition unrelated to changes in their visual attention (e.g., Hermer-Vazquez et al., 2001; Pyers et al., 2010). The weak verbal encoding hypothesis suggests that language and visual attention both facilitate children’s spatial performance but through different mechanisms (e.g., Dessalegn & Landau, 2008). The attentional hypothesis suggests that language only relates to spatial performance in that it is a useful cue in directing visual attention (Miller & Simmering, 2018). A limitation in current understanding of the mechanisms is that previous research has not measured children’s visual attention while performing spatial tasks under verbal versus visual conditions. The current study used eye-tracking to begin to differentiate among these hypotheses. In the study, 4- to 5-year-old children wore a head-mounted eye-tracker while participating in two spatial recall tasks, which assessed their use of an intrinsic reference frame during recall (memory for relations among nearby objects). Children first participated in the baseline recall task assessing their spontaneous use of an intrinsic reference frame and then in the cue manipulation recall task assessing performance under verbal, visual, or no cues (control).

This dissertation had three primary goals targeted to differentiating among the hypotheses: a) understand the effects of the conditions on recall accuracy; b) understand the effects of the conditions on eye-gaze patterns; and c) understand the effects of conditions and eye-gaze patterns on recall accuracy. First, the results showed that the conditions influenced children’s recall accuracy. Specifically, relative to baseline performance, the verbal and visual conditions facilitated children's recall performance, but no facilitation effect was found for children in the control condition. Comparing across conditions, children performed best in the verbal condition, and worst in the visual and control conditions, with differences between the visual and control conditions failing to reach significance. These results based only on recall performance are consistent with past research showing that children's spatial performance was most supported by verbal cues, with visual cues only minimally supporting performance (Miller et al., 2016).

Second, the results showed that the different conditions influenced children's eye-gaze patterns. Specifically, both the verbal and visual conditions increased children's fixations on the target landmarks during the hiding events relative to baseline, with no difference for participants in the control condition. Comparing across conditions, children fixated most on the target landmarks in the visual condition, followed by the verbal condition, with the least amount of target landmark fixations in the control condition. These results show that the visual cues were effective in directing children's eye-gaze. Inconsistent with the attentional hypothesis predictions, differences across the verbal and visual conditions did not result solely from the verbal cues directing children's visual attention more than the visual cues.

Third, the results showed that the conditions and children's eye-gaze patterns both uniquely predicted recall accuracy. Across all conditions, children who had more target landmark fixations tended to have higher recall accuracy. For the verbal condition, the verbal cues uniquely influenced children's recall accuracy irrespective of their eye-gaze patterns. However, there was also a mediated effect of eye-gaze such that hearing verbal cues directed children's gaze onto the target landmarks, in turn influencing their recall performance. For the visual condition, there was no unique effect of condition on recall accuracy not accounted for by visual attention. Specifically, the visual cues directed children's eye-gaze onto the target landmarks, which in turn influenced their recall accuracy. For the control condition there was no significant change in children's recall accuracy nor in their fixations on the target landmarks. These results are inconsistent with the strong verbal encoding hypothesis as children's visual attention influenced their recall accuracy and part of the effect of language was due to the influence of language on directing children's visual attention. Additionally, these results are also inconsistent with the attentional hypothesis as the effects of verbal cues on recall accuracy did not only result from changes in eye-gaze but were attributed to direct effects of verbal cues relating to children's spatial performance. Thus, the results are most in line with the weak verbal encoding hypothesis suggesting that spatial performance can be influenced by visual attention, but that language also has a unique effect possibly due to the format of representation (Dessalegn & Landau, 2008).

This study adds to the literature in three important ways. First, it is the first known study to directly measure children's visual attention to predict their spatial performance. Second, it is the first study to show that visual attention can be directed by both verbal and visual cues to support children's spatial performance. Third, it also provided direct evidence that verbal cues supported children's spatial performance independent of the effects of visual attention. To understand the theoretical implications of these findings, it is important to understand how they align with past research. As reviewed in the Introduction, prior research has used three major approaches to investigating the role of language in spatial cognition: cross-linguistic, experimental, and individual difference studies. Cross-linguistic studies showed that the ways in which one's language encodes spatial information can impact their spatial representations (e.g., Levinson, 2003). Additionally, individuals whose language lacks or has reduced consistency in using spatial words like "left" and "right", tend to perform poorly on spatial tasks requiring representation of such spatial relations (Gentner et al., 2013; Pyers et al., 2010). Experimental studies consistently showed benefits of verbal cues (both spatial and non-spatial) on children's spatial performance (e.g., Loewenstein & Gentner, 2005; Shusterman et al., 2011), with little if any benefit from non-verbal cues (Dessalegn & Landau, 2008; Miller et al., 2016). Correlational studies showed that children's spatial word usage predicted their spatial performance (e.g., Pruden et al., 2011). Further, some research proposed that children's general abilities to attend to task-relevant information underlay both spatial performance and language use, which then accounted for relations between spatial skills and language (Miller & Simmering, 2018).

The results from this dissertation align with effects found in past research, providing consistent evidence that both language and visual attention predict spatial performance. The results most closely parallel the experimental findings, conceptually replicating the finding that verbal cues benefit spatial performance more than visual cues (Dessalegn & Landau, 2008; Miller et al., 2016). However, this dissertation adds to the literature by showing that visual attention is important in spatial performance, even if visual manipulations are not as effective as verbal manipulations. This dissertation also aligns with some evidence from cross-linguistic and correlational studies. The current results support the notion from cross-linguistic studies that spatial language is highly beneficial in supporting spatial cognition and may provide a unique code for representing information that is not present through vision (e.g., Pyers et al., 2010). This is evident through the direct effects of the verbal cues condition on recall performance not accounted for by the mediated effect of landmark fixations. However, it remains an open question as to whether one could adopt a visual strategy to the extent of a linguistic strategy. Similar to individual difference studies (Miller & Simmering, 2018), this dissertation showed that visual attention predicted individual variability in spatial performance Specifically, this dissertation found strong relations between landmark fixations and recall accuracy across all conditions. Additionally, landmark fixations mediated relations with recall performance for both the verbal and visual cue condition, showing that changes in visual attention can augment, to some extent, spatial cognition. However, exactly why verbal cues were more effective than visual cues in supporting spatial performance, and whether effects of visual attention are truly nonlinguistic, are remaining open questions (addressed further below).

The current findings also parallel recent findings outside of spatial cognition investigating the effects of gesture and speech instruction on children's math equivalency problem solving (Wakefield, Novack, Congdon, Franconeri, & Goldin-Meadow, 2018). Specifically, when speech and gesture were provided during instruction, children's eye gaze followed along with the strategy being taught more than when only speech was provided. However, individual differences in eye gaze were only predictive of posttest performance when instruction was taught with both speech and gesture, not when it was only taught with speech. This suggests that gesture did not benefit learning just by directing children's attention. There was something unique about visually following along with the presence of gesture that did not occur with just visually following along with speech only. This effect is similar to the current study's findings that language can direct visual attention but language also does something more to support spatial performance.

In the sections that follow, I will discuss remaining open questions regarding why verbal cues provided by an experimenter were more effective than visual cues in supporting children's spatial performance; why individual differences in children's visual attention predict spatial performance when visual conditions are minimally effective; and then conclude with insights into mechanisms of developmental change and directions for future research.

Effectiveness of Verbal versus Visual Conditions

The current findings leave open questions regarding what was unique about the experimenter provided verbal cues that enhanced children's spatial performance relative to the visual cues. In the recall task, children needed to represent relations among the target cup and landmark(s) on the array, either through representing the target cup as "between" two landmarks (e.g., between the bear and cow) or representing the relation directionally relative to one landmark (e.g., in front of the pig, to the left of the frog). Language provides a coherent way of representing such relations, as a single word or set of words concretely conveys such spatial relations. However, it is unclear exactly how such relations are represented visually. Directing visual attention from one landmark to the other landmark does not necessary convey the same coherent information as using the word "between." As reviewed below, it is possible that language is providing a unique code for representing such relations that cannot be conveyed visually or to the same extent visually. Alternatively, it is possible that language is not special in terms of providing a particular type of representation but in the context of the current study was more effective due to its ability to provide pragmatic information, promoting effective instruction.

Language may provide children with a spatial representation that they cannot achieve non-verbally. For example, Gentner (2003) proposed a "relational shift" in children's reasoning resulting specifically from relational language. Accordingly, relational language drives children's representations of relations within the world, which are not obvious or salient without such language. When children do not have access to relational language (i.e., an inability to produce such words), they tend to focus more on concrete features or objects and fail to attend to relations among them. The current analyses were focused on fixations to the target cups versus target landmarks, which are both concrete objects. It is likely that children who knew to fixate on the landmarks were processing relational information, as the landmarks are only relevant for the task based on how their locations are relative to the target cup. Attention to the relation would not be evident through the current measures (either number or duration of fixations to landmarks). However, it is possible that another measure, such as the frequency of children's gaze switches between the cup and landmark, which I can investigate in future analyses could account for additional variance that corresponds to children's representation of the relational information. In the verbal condition, the language provided by the experimenter likely enabled children to form such relational representation at encoding, more than children would have formed on their own without using such relational language.

Alternatively, in line with the weak verbal encoding hypothesis, it may be that language temporarily augments cognition. Language can provide an abstract representation separate from vision that facilitates the encoding and maintenance of spatial information (e.g., Dessalegn & Landau, 2008). In the current study, the abstract language "between" may have given children a coherent way of representing the spatial information that augmented their visual representation of such information. Unlike the relational shift hypothesis, it is not that children could not form such relations visually, as seen in the associations between children's fixations on target landmarks and their probability correct. However, it suggests that the language strengthened their representations beyond what was provided through vision. In the visual condition, the added visual cues did not necessarily provide children with a more coherent way of representing the information. For instance, the visual cues may have had children attend to the landmarks more, but not necessarily to the relation between the landmark and the target cup.

Similarly, children often learn most effectively when they receive information through multiple modalities. For instance, research has shown that children benefit from representing new information through both pictures and words (e.g., Mayer, 2002). Additionally, children learning math problems benefit most when the instruction is presented through both speech and gesture than from just speech alone (Congdon et al., 2017; Wakefield et al., 2018). In the context of the current study it is unknown whether pairing the visual cues with verbal cues would have enhanced children's representation even more than verbal cues alone (parallel to findings that gesture and speech together are better than either on its own). However in the current study, children received a form of visual information from the array itself. The verbal condition provided children with a different modality of representing the spatial information, while the visual condition provided enhanced cues but within the same modality. Irrespective of whether language itself is providing distinct information from vision, representing information across modalities can facilitate children's learning. Representing spatial information through multiple modalities may have also reduced children's need to represent the information as thoroughly through vision, reallocating the cognitive resources that children used.

Beyond possible effects of providing a different code of representation, verbal cues may have provided children with more pragmatic information facilitating their learning. Children are frequently instructed through language and thus language is a naturalistic means by which children receive instruction. There may be nothing unique about the specific linguistic representation but instead language is unique in providing pragmatic information. For example, Shustman et al. (2011) demonstrated that relational language was not the only type of language that could facilitate children's spatial performance in a disorientation search task. Rather, any language that highlighted the pragmatic relevance of cues was useful. In current study, the instructions from the experimenter before each recall task were intended to highlight the pragmatic relevance of the cues (i.e., in the verbal condition: "listen to what I say"; in the visual condition: "look at what I am doing”). However, due to language providing a more naturalistic means of communicating relevant information, it may have been easier for children to understand that the verbal cues were providing relevant information compared to the visual cues. Shaking the landmarks to represent “between” without using language does not convey the same pragmatic information as using the word “between.” In a previous version of this study, the experimenter moved the landmarks on top of the cup and back down to show the spatial relation, which likely highlighted the spatial relation more, but this manipulation was also significantly less effective than the verbal manipulation.

It is often less natural and more difficult to provide pragmatic information non-verbally about the relevance of relational information than providing the same information verbally. However, there may be special cases, not employed in the current study in which non-verbal cues provide more pragmatic information and are as effective as verbal cues. For example, gesture often occurs in communicative contexts and can provide relevant information with or without accompanying speech. It would be informative in future research to test effects of gesture versus verbal cues on children’s spatial performance and investigate if the verbal cues benefits remain. One study tested this effect by assessing children’s performance on a puzzle assembly task (Young, Cartmill, Levine, & Goldin-Meadow, 2014). The research found that children performed best under conditions of speech and gesture (dual representation) than with conditions of speech alone, gesture alone, or neither. However with the nature of their task, the spatial language on its own was not an effective cue making it difficult to compare effects of gesture alone versus speech alone.

The visual cues manipulation may have also been too salient or distracting for the child. First, successful visual manipulations with adults have typically involved subtle manipulations where the adult has little awareness that their eye-gaze is being directed in a particular manner (e.g., Bailey et al., 2012; Grant & Spivey, 2003). Similarly, children struggle to learn when the format of the information provided is too salient. For example, when using manipulatives to learn math problems, children can struggle to learn with perceptually rich manipulatives (McNeil, Uttal, Jarvin, & Sternberg, 2009). In the current study's visual condition relative to baseline, children swapped their fixations from primarily fixating on the target cup to fixating on the target landmarks. These increased fixations onto the target landmark demonstrate that the visual condition was very salient. However, the salience of this visual manipulation may have distracted the child, overly centering their focus on the landmarks, inhibiting their attention to the relevant relational information between the cups and landmarks. Relatedly, children may have been focused on the shaking motion and not on the fact that the shaking is revealing something special about the property of the landmarks and their relation to the target cup. Thus, the extra-salience may have inadvertently deterred children from using the landmarks. In sum, verbal cues are an effective means for guiding children within spatial tasks. This results possibly in the ways in which the cues lead children to form a new representation and/or because they provide pragmatic didactic cues and keep children on task, helping them identify relevance components. It continues to be unknown whether visual cues can ever be as effective as verbal cues in supporting children's spatial performance.

Children's Spontaneous Visual Attention

There is undoubtedly clear experimental evidence that verbal cues are more effective than visual cues in supporting spatial performance. However, then why have studies (Miller & Simmering, 2018), including the current study, shown that individual differences in children's visual attention relate to systematic variations in their spatial performance irrespective of effects of condition? Additionally, while a relatively small effect, there were improvements in the visual condition due to the effects of the condition in directing eye-gaze. It is possible that the effects found with eye-gaze were just a measure of language use—children who fixate more on the landmarks are also verbally encoding the information into language more. Alternatively, it is possible that the measure of eye-gaze reflects basic attentional mechanisms—children who direct their attention better, irrespective of using language, perform better on spatial tasks.

Considering that verbal cue manipulations are very effective in supporting children's spatial performance, individual differences in children's eye-gaze patterns may not specifically reflect the use of a visual strategy but may have arisen through young children's use of verbal encoding. If using verbal coding, children may explicitly adopt a verbal strategy, actively encoding and rehearsing information on the array in language or their language knowledge may implicitly direct them to attend to relevant information. If children spontaneously used language explicitly or implicitly to support their recall performance, they would need to have some fixations onto the target landmarks. There may be fewer fixations on the target landmarks when using such a verbal encoding strategy relative to a visual strategy as children would have a second code of representing such information; but there would still need to be some fixations to know what objects to encode. As reviewed in the Introduction there is not strong evidence that children or adults spontaneously use verbal encoding strategies in spatial tasks (Nardini et al., 2006; Ratliff & Newcombe, 2008), but there is strong evidence with adults that their language knowledge shapes their spatial representations (e.g., Choi & Hattrup, 2012; Levinson, 2003). Similarly, there is evidence that both children and adults do not always use their knowledge of language spontaneously unless it is activated (Feist & Gentner, 2007; Scott & Sera, 2018). While evidence remains mixed, it is important to acknowledge that it is unknown the extent to which verbal encoding was used across conditions. Relatedly, it is unknown the extent to which children with knowledge of relevant spatial words may have encoded visual information differently based on their linguistic knowledge.

Understanding whether children's eye-gaze patterns reflect a form of explicit or implicit verbal encoding is a difficult question to answer. As discussed in the section above, it is unknown exactly what an eye-gaze pattern associated with a particular spatial word would look like. Additionally, there are multiple different ways in which one could verbally encode relevant relations on the array and so any one particular pattern may or may not reflect children's use of verbal encoding. There is some work with adults investigating eye-gaze patterns while adults make judgments of spatial words, where there is a tendency to fixate in a particular order (Coventry et al., 2010; Franconeri, Scimeca, Roth, Helseth, & Kahn, 2012). Beyond this dissertation, I plan to continue investigating this question to understand whether individual differences in children's eye-gaze patterns may reflect their use of verbal encoding. I will analyze data from my production task to see whether there are consistencies in children's eye-gaze patterns when they produce particular types of spatial words and other task-relevant information. Additionally, I will assess whether there are similarities in eye-gaze patterns when producing relevant information during the production task and when encoding information on the array during the recall tasks. While indirect evidence, this type of analysis may provide some insight and directions for future research into the potential of children's language knowledge influencing their spatial representations.

Alternatively, visual attention may be underlying at least some of the effects related to developmental change in spatial cognition. As discussed in the section above, the weak effect of the visual condition relative to the verbal condition could have resulted from a variety of factors. The difference between how the experimental manipulation versus individual differences in eye-gaze predicted spatial performance may reflect differences in how children processed and attended to the information on the array. Attention can be endogenously controlled through voluntary shifts or it can be exogenously controlled through automatic non-voluntary shifts (Theeuwes, 1994). Children's fixations on the target landmarks in the baseline and control conditions reflect endogenous attention, as no specific cues directed their attention to that information. Additionally, fixations on the target landmarks in the verbal condition reflect endogenous attention as children would voluntarily switch their visual attention to the specific landmarks after hearing the cues. In contrast, the fixations in the visual condition more likely reflect exogenous attention, driven by the experimenters' movement of the objects. These differences in the source of attentional control may relate to why individual differences in attention predict recall accuracy, despite the relative ineffectiveness of the visual cues in supporting search performance.

A central question is what factors were driving children's endogenous attention to the target landmarks in the baseline and control conditions. Similar to how the verbal cues directed attention in the verbal condition, it could be that children's language knowledge in the baseline and control conditions directed their attention to the landmarks (e.g., Loewenstein & Gentner, 2005). However, studies have shown that children's knowledge of spatial words does not reliably direct them to endogenously attend to the relevant pieces of information (Miller & Simmering, 2018; Miller et al., 2017; Scott & Sera, 2018). Accordingly, children's knowledge of the word "between" in the baseline and control conditions may not have directed them to attend to the landmarks that were relevant for encoding a between relation. Instead children needed to endogenously attend to this relation first to recognize the potential relevance of the word "between." Thus it is possible that individual differences in landmark fixations predicting recall accuracy reflected children's non-verbal attention, even though verbal cues were effective in supporting children's spatial performance.

Mechanisms of Developmental Change

These issues related to differences in effects of experimenter provided cues versus children's visual attention raise larger questions regarding the extent to which the effects of experimental manipulations truly reflect change processes in cognition. On the one hand, there are developmental changes in the effectiveness of verbal cue manipulation, whereby older children do not benefit from the additional cues – there are no differences between children in a verbal versus control condition (Dessalegn & Landau, 2013; Loewenstein & Gentner, 2005). It is often assumed that the older children do not need language support because they are already using a verbal encoding strategy. However, on the other hand, high performance in a control condition does not necessarily reflect children's use of a verbal strategy; children could be adopting other effective non-verbal strategies that equate performance in the verbal condition. Additionally, younger children can benefit from verbal cues because they are effective cues in communicating relevant information even if that is not a strategy that children necessarily spontaneously adopt and use later in development.

What can be concluded about mechanisms of developmental change from the current study? In line with the weak verbal encoding hypothesis, we can conclude that there are changes in children's visual attention over development that relate to spatial performance differences. Young children can easily adopt verbal strategies when provided with didactic support, as language can provide a different means of representing spatial information. However, open questions remain as to the extent to which changes in children's use of verbal encoding is necessarily causing changes in their spatial cognition. In terms of avenues of future research, we can continue to refine the visual condition and investigate whether such conditions can be more effective. For example, we could adopt a manipulation using more pragmatic visual cues such as gesture. Additionally, we could conduct a yoked design where we match high performing children from the control and baseline conditions patterns of eye-gaze and try to implicitly direct children's eye-gaze in the same manner (cf. Bailey et al., 2012). If these manipulations are more effective, it would provide stronger evidence that non-verbal visual attention can support spatial performance and possibly influence developmental change.

However, in order to truly understand developmental change in spatial cognition, we need to move research beyond assessing the effects of one-visit experimental and correlational studies and conduct in-depth longitudinal investigations with multiple measures of real-time processing. One example would be to conduct a longitudinal study with multiple visits during the age period where there are significant changes in children's spatial word comprehension and production. During the visits, one could measure children's eye-gaze while performing spatial tasks and take in-depth language measures including measures of task-relevant language use and spatial vocabulary. Over time, this type of study would assess whether changes in eye-gaze during spatial tasks precede or follow changes in children's spatial word and task-relevant language use. Additionally, research should use other implicit measures of cognitive processing in real-time. For example, one could measure neurological activity in both language areas of the brain such as Broca's or Wernicke's area (Damasio & Damasio, 1992) and areas thought to be involved in spatial processing such as the hippocampus (Morris, Garrud, Rawlins, & O'Keefe, 1982). There is some experimental evidence that block play training in 8-year-old children results in changes in brain areas associated with spatial activation, specifically the anterior lobe of the cerebellum, right parahippocampus, and bilateral fusion gyrus. One could extend this type of research and measure changes in activation in these different brain regions across high and low performing children and assess similarity and differences in activation when provided with verbal cue support.

In actuality, developmental change in spatial cognition is likely not caused by one primary mechanism but by multiple mechanisms interacting over development. Language and visual attention are likely not redundant processes but interact with each other over development to support spatial cognition. In children's everyday life as they hear spatial language, they are likely not only learning to attend to relevant pieces of information but begin to develop a strong linguistic representation that moves beyond what is provided solely from vision. As children explore their environment, engaging in spatial activities, they likely learn to attend to what is relevant and use this attention to further support their language use and their spatial understanding. In conclusion, this dissertation provided a more in-depth understanding of processes underlying spatial cognition, demonstrating that both visual attention and language processes relate in unique ways to support young children's spatial performance. Future research should continue to use implicit measures and assess children at multiple time points to better understand mechanisms underlying developmental change.

Acknowledgments

Thank you to my mentor Dr. Vanessa Simmering for providing me with guidance through my dissertation as well as through my undergraduate thesis and graduate career. She has been extremely supportive in helping me to create a programmatic line of research with my interest in children's language use and her expertise in spatial cognition and memory development. I cherish the many hours that she has spent meeting with me and thinking deeply and critically about my research. She has helped me through all steps of the research process, shaping me to become a strong researcher for a career in academia.

Thank you to Dr. Heather Kirkorian and Seung Heon Yoo for providing extensive training on eye-tracking procedures. Dr. Kirkorian was very influential in ensuring that my experimental design was suitable for head mounted eye-tracking and for training me on best practices for handling eye-tracking data. She created a Python code for the eye-tracking variables specific for my dissertation. Seung was instrumental in helping to pilot my dissertation data, training me on how to use the eye-tracking equipment and software, and collecting data for me as needed.

Thank you to my other committee members Dr. Jenny Saffran, Dr. Martha Alibali, and Dr. Haley Vlach for providing careful and thoughtful advice in helping me design my dissertation study. All three members have been very influential in supporting me throughout my graduate career.

Thank you to Mark Banghart, a statistical consultant at the Social Science Computing Cooperative at the University of Wisconsin for the extensive training and advising on conducting appropriate statistics using mixed effects models. Thank you to the ample undergraduate research assistants in both the Space Lab and the Cognitive Development Media Lab for helping me collect my dissertation data. The undergraduates in the Space Lab were also extremely influential in coding my data (spending 30+ hours a week for 2 semesters). Additionally, thank you to all the families in the Madison area who came into the lab to participate in my dissertation.

Funding for participant recruitment was provided by Waisman Intellectual and Developmental Disabilities Research Center grant (P30HD03352). Financial support was provided to me through the Vision Research Trainee Grant, sponsored by the University of Wisconsin’s McPherson Eye-Research Institution and from the Marian A. Schwartz Research Fellowship, sponsored by the University of Wisconsin Psychology Department.

References

Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypothesis. *Psychological Methods*, *2*(4), 315–328. https://doi.org/10.1037/1082-989X.2.4.315

Andersen, N. E., Dahmani, L., Konishi, K., & Bohbot, V. D. (2012). Eye tracking, strategies, and sex differences in virtual navigation. *Neurobiology of Learning and Memory*, *97*(1), 81–89. https://doi.org/10.1016/j.nlm.2011.09.007

Axelsson, E. L., Perry, L. K., Scott, E. J., & Horst, J. S. (2016). Near or far: The effect of spatial distance and vocabulary knowledge on word learning. *Acta Psychologica*, *163*, 81–87. https://doi.org/10.1016/j.actpsy.2015.11.006

Bailey, R., McNamara, A., Costello, A., Sridharan, S., & Grimm, C. (2012). Impact of subtle gaze direction on short-term spatial information recall. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 67–74). ACM.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

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

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. *Language in Mind: Advances in the Study of Language and Cognition*, 387–428.

Bremner, J. G. (1978). Egocentric versus allocentric spatial coding in nine-month-old infants: Factors influencing the choice of code. *Developmental Psychology*, *14*, 346–355.

Bunger, A., Trueswell, J. C., & Papafragou, A. (2012). The relation between event apprehension and utterance formulation in children: Evidence from linguistic omissions. *Cognition*, *122*(2), 135–149. https://doi.org/10.1016/j.cognition.2011.10.002

Chabal, S., Schroeder, S. R., & Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Attention, Perception, & Psychophysics*, *77*(8), 2684–2693. https://doi.org/10.3758/s13414-015-0973-7

Choi, S., & Hattrup, K. (2012). Relative contribution of perception/cognition and language on spatial categorization. *Cognitive Science*, *36*(1), 102–129. https://doi.org/10.1111/j.1551-6709.2011.01201.x

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. In *Advances in Child Development and Behavior* (Vol. 34, pp. 283–322). Elsevier. https://doi.org/10.1016/S0065-2407(06)80010-8

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O’Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, *50*, 65–74. https://doi.org/10.1016/j.learninstruc.2017.03.005

Coventry, K. R., Lynott, D., Cangelosi, A., Monrouxe, L., Joyce, D., & Richardson, D. C. (2010). Spatial language, visual attention, and perceptual simulation. *Brain and Language*, *112*(3), 202–213. https://doi.org/10.1016/j.bandl.2009.06.001

Damasio, A. R., & Damasio, H. (1992). Brain and Language. *Scientific American*, *267*(3), 88–109.

Dessalegn, B., & Landau, B. (2008). More than meets the eye: the role of language in binding and maintaining feature conjunctions. *Psychological Science*, *19*, 189–195. https://doi.org/10.1111/j.1467-9280.2008.02066.x

Dessalegn, B., & Landau, B. (2013). Interaction between language and vision: it’s momentary, abstract, and it develops. *Cognition*, *127*, 331–344. https://doi.org/10.1016/j.cognition.2013.02.003

Farran, E. K., & O’Leary, B. (2016). Children’s ability to bind and maintain colour–location conjunctions: the effect of spatial language cues. *Journal of Cognitive Psychology*, *28*(1), 44–51. https://doi.org/10.1080/20445911.2015.1092980

Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, *35*(2), 283–296. https://doi.org/10.3758/BF03193449

Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. E. (2012). Flexible visual processing of spatial relationships. *Cognition*, *122*(2), 210–227. https://doi.org/10.1016/j.cognition.2011.11.002

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, *44*(03), 677–694. https://doi.org/10.1017/S0305000916000209

French, R. M., & Thibaut, J.-P. (2014). Using eye-tracking to predict children’s success or failure on analogy tasks. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Quebec City, CA.: Cognitive Science Society.

Gentner, D. (2003). Why we’re so smart. In D. Gentner & S. Golden-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.

Gentner, D., Özyürek, A., Gürcanli, Ö., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition*, *127*(3), 318–330.

Glady, Y., Thibaut, J.-P., French, R. M., & Blaye, A. (2012). Examining children's failure in analogy making tasks: A problem of focus of attention? In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Sapporo, Japan: Cognitive Science Society.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: guiding attention guides thought. *Psychological Science*, *14*(5), 462–466. https://doi.org/10.1111/1467-9280.02454

Haendler, Y., Kliegl, R., & Adani, F. (2015). The impact of pronoun type and grammatical skills on children's processing of object relative clauses. In *Proceedings of the 39th Annual Boston University Conference on Language Development*. Cascadilla Press Somerville, MA.

Hegarty, M., & Waller, D. A. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 121–169). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511610448.005

Hermer-Vazquez, L., Moffet, A., & Munkholm, P. (2001). Language, space, and the development of cognitive flexibility in humans: the case of two spatial memory tasks. *Cognition*, *79*, 263–299. https://doi.org/10.1016/S0010-0277(00)00120-7

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, *39*, 3–36. https://doi.org/10.1006/cogp.1998.0713

Hitch, G. J., Halliday, S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition*, *16*(2), 120–132.

Hoffman, J. E., Landau, B., & Pagani, B. (2003). Spatial breakdown in spatial construction: Evidence from eye fixations in children with Williams syndrome. *Cognitive Psychology*, *46*, 260–301.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50*(3), 346–363.

Hund, A. M., Bianchi, L. J., Winner, J. F., & Hesson-McInnis, M. S. (2017). Complex spatial language improves from 3 to 5 years: The role of prompting and overhearing in facilitating direction giving using between and middle. *Cognitive Development*, *43*, 170–181. https://doi.org/10.1016/j.cogdev.2017.04.002

Jankowski, J. J., Rose, S. A., & Feldman, J. F. (2001). Modifying the distribution of attention in infants. *Child Development*, *72*(2), 339–351. https://doi.org/10.1111/1467-8624.00282

Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience*, *29*(7), 877–892. https://doi.org/10.1080/01690965.2013.804941

Jensen, C., Miller, H. E., & Simmering, V. R. (2015, March). Developmental changes in how children weight spatial information in recall. Poster presented at the The 71st Biennial Meeting of the Society for Research in Child Development, Philadelphia, PA.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kastens, K. A., & Ishikawa, T. (2006). Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields. In *Special Paper 413: Earth and Mind: How Geologists Think and Learn about the Earth* (Vol. 413, pp. 53–76). Geological Society of America. https://doi.org/10.1130/2006.2413(05)

Kenward, M. G., & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, *53*(3), 983. https://doi.org/10.2307/2533558

Kirkorian, H. L., Lavigne, H. J., Hanson, K. G., Troseth, G. L., Demers, L. B., & Anderson, D. R. (2016). Video deficit in toddlers' object retrieval: What eye movements reveal about online cognition. *Infancy*, *21*(1), 37–64. https://doi.org/10.1111/infa.12102

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Landau, B., Dessalegn, B., & Goldberg, A. M. (2010). Language and space: Momentary interactions. In P. Chilton & V. Evans (Eds.), *Language, cognition, and space: The state of the art and new directions.* (pp. 51–78). London: Equinox Publishing.

Lauer, J. E., & Lourenco, S. F. (2016). Spatial processing in infancy predicts both spatial and mathematical aptitude in childhood. *Psychological Science*, *27*(10), 1291–1298. https://doi.org/10.1177/0956797616655977

Learmonth, A. E., Nadel, L., & Newcombe, N. S. (2002). Children’s use of landmarks: Implications for modularity theory. *Psychological Science*, *13*, 337–341. https://doi.org/10.1111/j.0956-7976.2002.00461.x

Learmonth, A. E., Newcombe, N. S., & Huttenlocher, J. (2001). Toddlers’ use of metric information and landmarks to reorient. *Journal of Experimental Child Psychology, 80*, 225–244. https://doi.org/10.1006/jecp.2001.2635

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*, 940. https://doi.org/10.1037/0012-1649.35.4.940

Levinson, S. C. (2001). Covariation between spatial language and cognition, and its implications for language learning. *Language Acquisition and Conceptual Development*, 566–588.

Levinson, S. C. (2003). *Space in language and cognition: explorations in cognitive diversity*. Cambridge ; New York: Cambridge University Press.

Li, P., Abarbanell, L., Gleitman, L., & Papafragou, A. (2011). Spatial reasoning in Tenejapan Mayans. *Cognition*, *120*(1), 33–53. https://doi.org/10.1016/j.cognition.2011.02.012

Li, X., Carlson, L. A., Mou, W., Williams, M. R., & Miller, J. E. (2011). Describing spatial locations from perception and memory: The influence of intrinsic axes on reference object selection. *Journal of Memory and Language*, *65*(2), 222–236. https://doi.org/10.1016/j.jml.2011.04.001

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*, 315–353. https://doi.org/10.1016/j.cogpsych.2004.09.004

Mayer, R. E. (2002). Multimedia learning. In *Psychology of Learning and Motivation* (Vol. 41, pp. 85–139). Elsevier. https://doi.org/10.1016/S0079-7421(02)80005-6

McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction*, *19*(2), 171–184. https://doi.org/10.1016/j.learninstruc.2008.03.005

Miller, H. E., Patterson, A. R., & Simmering, V. R. (2016). Language supports young children's use of spatial relations to remember locations. *Cognition*, *150*, 170–180. https://doi.org/doi:10.1016/j.cognition.2016.02.006

Miller, H. E., & Simmering, V. R. (2018). Children's attention to task-relevant information accounts for relations between language and spatial cognition. *Journal of Experimental Child Psychology*, *172*, 107–209. https://doi.org/10.1016/j.jecp.2018.02.006

Miller, H. E., Vlach, H. A., & Simmering, V. R. (2017). Producing spatial words is not enough: Understanding the relation between language and spatial cognition. *Child Development*. https://doi.org/10.1111/cdev.12664

Miura, I. T., Kim, C. C., Chang, C., & Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons. *Child Development*, 1445–1450. https://doi.org/10.2307/1130659

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., & O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, *297*(5868), 681–683. https://doi.org/10.1038/297681a0

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 162–170. https://doi.org/10.1037//0278-7393.28.1.162

Nardini, M., Burgess, N., Breckenridge, K., & Atkinson, J. (2006). Differential developmental trajectories for egocentric, environmental and intrinsic frames of reference in spatial memory. *Cognition*, *101*, 153–172. https://doi.org/10.1016/j.cognition.2005.09.005

Nazareth, A., Odean, R., & Pruden, S. M. (2017). The use of eye-tracking in spatial thinking research. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 239–260). IGI Global. https://doi.org/10.4018/978-1-5225-1005-5

Negen, J., & Nardini, M. (2015). Four-year-olds use a mixture of spatial reference frames. *PLOS ONE*, *10*(7), e0131984. https://doi.org/10.1371/journal.pone.0131984

Newcombe, N., Huttenlocher, J., Drummey, A. B., & Wiley, J. G. (1998). The development of spatial location coding: Place learning and dead reckoning in the second and third years. *Cognitive Development*, *13*, 185–200.

Ohm, C., Mueller, M., Ludwig, B., & Bienk, S. (2014, September). *Where is the landmark? Eye rracking studies in large-scale indoor environments*. Presented at the 2nd International Workshop on Eye Tracking for Spatial Research, Vienna, Austria.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language*, *74*(3), 557. https://doi.org/10.2307/417793

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*(2–3), 227–272. https://doi.org/10.1016/0001-6918(94)90004-3

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717–731. https://doi.org/10.3758/BF03206553

Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science*, *14*, 1417–1430. https://doi.org/10.1111/j.1467-7687.2011.01088.x

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 12116–12120.

Ratliff, K. R., & Newcombe, N. S. (2008). Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cognitive Psychology, 56*, 142–163. https://doi.org/10.1016/j.cogpsych.2007.06.002

Robinson, D. A. (1965). The mechanics of human smooth pursuit eye movement. *The Journal of Physiology*, *180*(3), 569–591. https://doi.org/10.1113/jphysiol.1965.sp007718

Scott, N. M., & Sera, M. D. (2018). Language unifies relational coding: The roles of label acquisition and accessibility in making flexible relational judgments. *Journal of Memory and Language*, *101*, 136–152. https://doi.org/10.1016/j.jml.2018.04.002

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects.

Shusterman, A., Lee, S., & Spelke, E. S. (2011). Cognitive effects of language on human navigation. *Cognition*, *120*, 186–201. https://doi.org/10.1016/j.cognition.2011.04.004

Shusterman, A., & Spelke, E. (2005). Language and the development of spatial reasoning. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and concents* (pp. 89–106). Oxford, UK: Oxford University Press.

Simms, N., & Gentner, D. (2008). Spatial language and landmark use: Can 3-, 4-, and 5-year-olds find the middle? In *29th Annual Conference of the Cognitive Science Society* (pp. 191–196). Washingtion, D.C.: Lawrence Erlbaum Associates.

Sorby, S. A. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education*, *31*(3), 459–480. https://doi.org/10.1080/09500690802595839

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, *12*(4), 282–286. https://doi.org/10.1111/1467-9280.00352

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4no label), 799–806.

Thibaut, J.-P., & French, R. M. (2016). Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking. *Cognitive Development*, *38*, 10–26. https://doi.org/10.1016/j.cogdev.2015.12.002

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, *14*(4), 663–668. https://doi.org/10.3758/BF03196818

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*, *59*(5), 1–38.

Vasilyeva, M., & Lourenco, S. F. (2010). Spatial development. In R. M. Lerner, M. E. Lamb, & A. M. Freund (Eds.), *The Handbook of Life-Span Development*. Hoboken, NJ, USA: John Wiley & Sons, Inc. Retrieved from http://doi.wiley.com/10.1002/9780470880166.hlsd001020

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). Links between spatial and math skills across the preschool years. *Monographs of the Society for Research in Child Development*, *82*(1), 71–80. https://doi.org/10.1111/mono.12283

Viaene, P. G. U.-G. D., Vansteenkiste, P. G. U., Lenoir, M. G. U., De Wulf, A. G. U., & De Maeyer, P. G. U. (2016). Examining the validity of the total dwell time of eye fixations to identify landmarks in a building. *Journal of Eye Movement Research*, *9*(3), 1–11. https://doi.org/10.16910/jemr.9.3.4

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*, 817–835. http://dx.doi.org/10.1037/a0016127

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science*, e12664. https://doi.org/10.1111/desc.12664

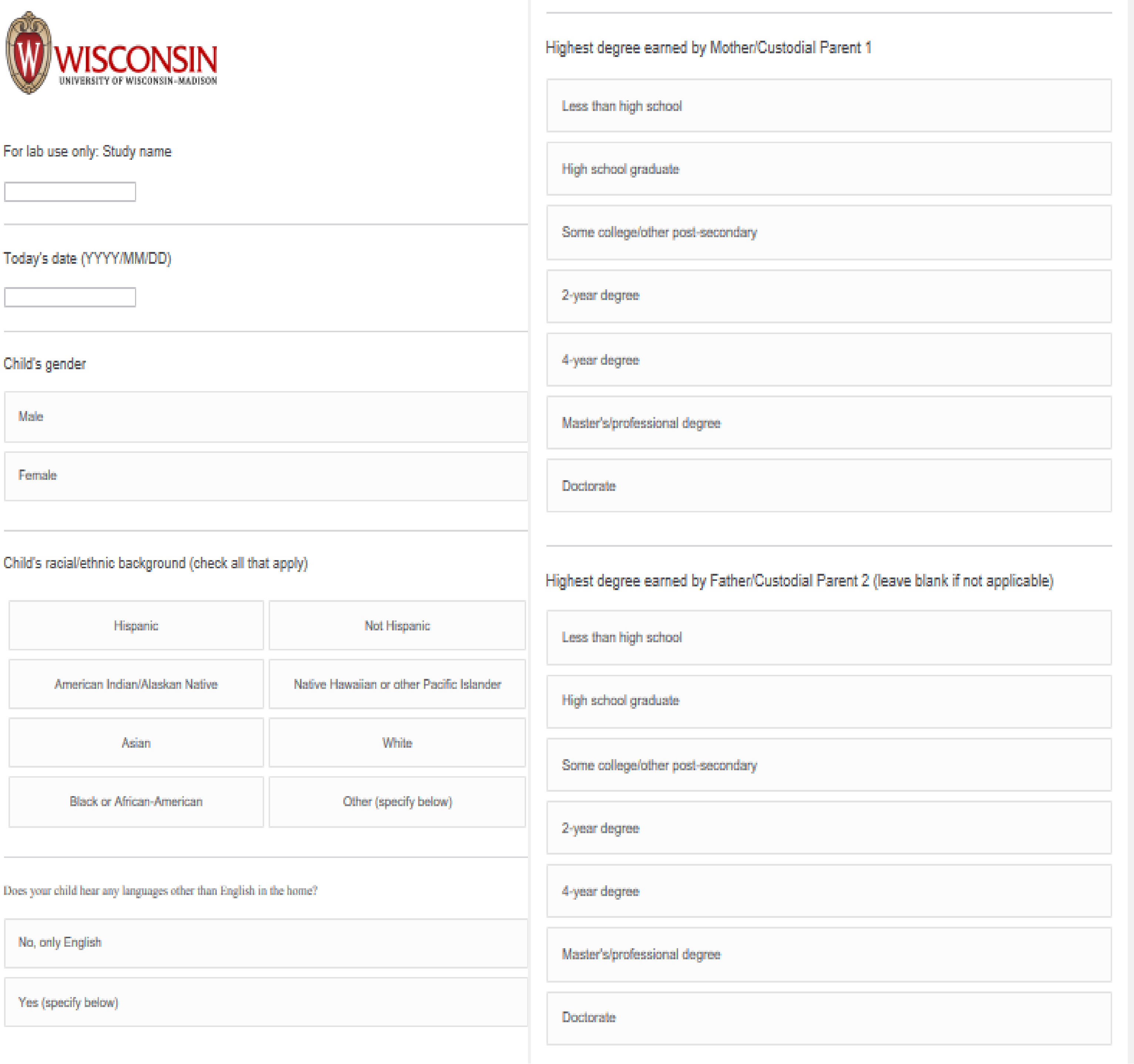
Wu, H., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492. https://doi.org/10.1002/sce.10126

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118–136. https://doi.org/10.1016/j.jecp.2010.04.014

Young, C. J., Cartmill, E., A. ., Levine, S. C., & Goldin-Meadow, S. (2014). Gesture and speech input are interlocking pieces: The development of children's jigsaw puzzle assembly ability. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Appendix A

A1. Demographics questionnaire administered through online qualtrics survey



For lab use only: Study name

Today's date (YYYY/MM/DD)

Child's gender

Male

Female

Child's racial/ethnic background (check all that apply)

Hispanic	Not Hispanic
American Indian/Alaskan Native	Native Hawaiian or other Pacific Islander
Asian	White
Black or African-American	Other (specify below)

Does your child hear any languages other than English in the home?

No, only English

Yes (specify below)

Highest degree earned by Mother/Custodial Parent 1

Less than high school

High school graduate

Some college/other post-secondary

2-year degree

4-year degree

Master's/professional degree

Doctorate

Highest degree earned by Father/Custodial Parent 2 (leave blank if not applicable)

Less than high school

High school graduate

Some college/other post-secondary

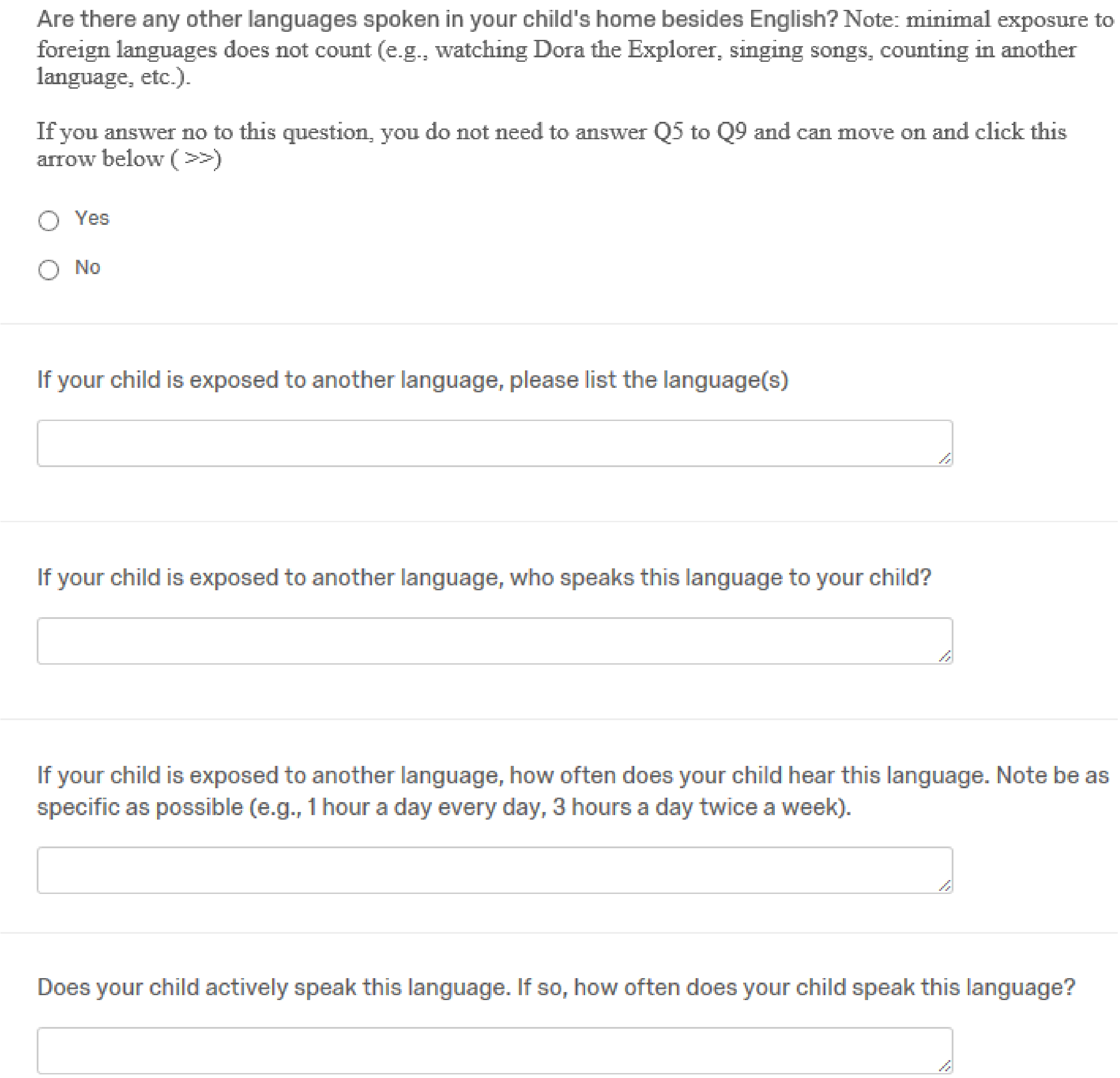
2-year degree

4-year degree

Master's/professional degree

Doctorate

A2. Language questionnaire administered through online Qualtrics survey



Are there any other languages spoken in your child's home besides English? Note: minimal exposure to foreign languages does not count (e.g., watching Dora the Explorer, singing songs, counting in another language, etc.).

If you answer no to this question, you do not need to answer Q5 to Q9 and can move on and click this arrow below (>>)

○ Yes

○ No

If your child is exposed to another language, please list the language(s)

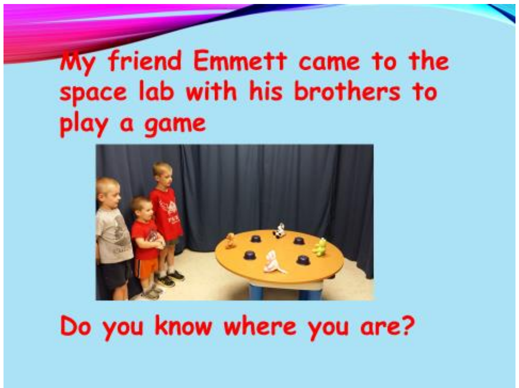
If your child is exposed to another language, who speaks this language to your child?

If your child is exposed to another language, how often does your child hear this language. Note be as specific as possible (e.g., 1 hour a day every day, 3 hours a day twice a week).

Does your child actively speak this language. If so, how often does your child speak this language?

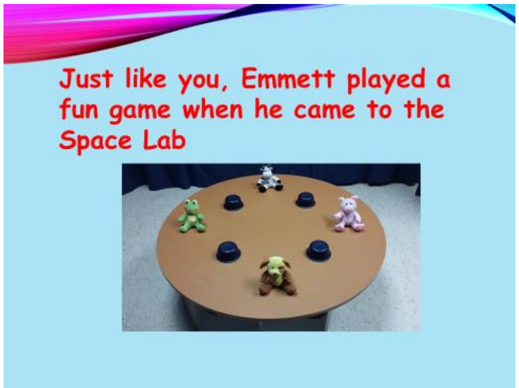
A3. Book to familiarize children with eye-tracking cap before the session.

My friend Emmett

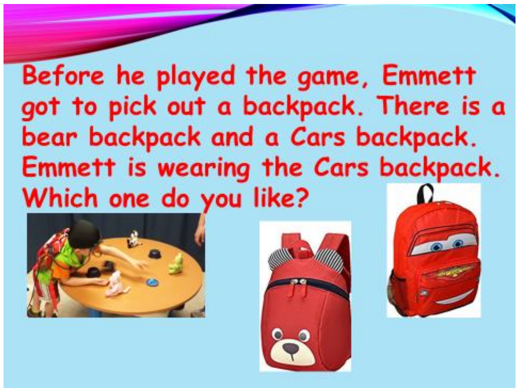


My friend Emmett came to the space lab with his brothers to play a game

Do you know where you are?



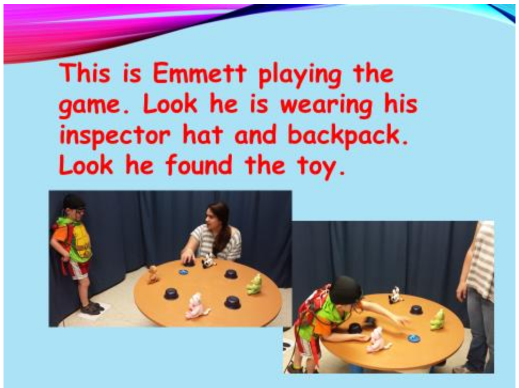
Just like you, Emmett played a fun game when he came to the Space Lab



Before he played the game, Emmett got to pick out a backpack. There is a bear backpack and a Cars backpack. Emmett is wearing the Cars backpack. Which one do you like?

Emmett also wore this special inspector hat. This hat helps him find toys.

It is just like inspector gadget's hat. Look they look the same.



This is Emmett playing the game. Look he is wearing his inspector hat and backpack. Look he found the toy.

Are you ready to play the game and be an inspector? Just like Emmett

Appendix B

Table B1

Trial exclusions due to eye-tracking quality issues

Exclusion criterion	Control		Verbal		Visual		All conditions	
	# trials	*Prop. trials*	*# trials*	*Prop. trials*	*# trials*	*Prop. trials*	*# trials*	*Prop. trials*
Total (inclusions and exclusions)	635	1.00	688	1.00	696	1.00	2019	1.00
No exclusions	491	.77	510	.74	583	.84	1584	.78
Exclusions	144	.23	178	.26	133	0.16	435	0.22
Exclusion by type								
No issue with these trials but participant had too many other exclusions	37	.06	50	*.07*	17	.02	104	.05
Calibration	4	.01	19	.03	12	.02	35	.02
No target cup fixation (distracted)	5	.01	3	.00	10	.01	18	.01
No target cup fixation (fixation criterion not reached)	7	.01	9	.01	12	.02	28	.01
Low % fixations captured	39	.06	50	.07	21	.03	110	.05
High deviant pupil diameter	13	.02	5	.01	13	.02	31	.02
Fixation coded as unreliable	10	.02	18	.03	17	.02	45	.02
First target cup fixation lost	7	.01	8	.01	6	.01	21	.01
Low % fixations captured; High deviant pupil diameter	12	.02	7	.01	4	.01	23	.01
Low % fixations captured; Fixation coded as unrelaible	6	.01	7	.01	1	<.01	14	.01
Low % fixations captured; First target cup fixation lost	1	.00	0	.00	0	.00	1	<.01
Low % fixations captured; High deviant pupil diameter; Fixation coded as unreliable	3	.00	2	.00	0	.00	5	<.01

Table B2

Data quality statistics with outlier trials included (except trials with poor calibration and no target cup fixations)

		Fixation duration (ms)		Trial duration (ms)		Fixation duration/trial duration		Good pupil frames		Poor eye-tracking fixations	
Recall Task	Condition	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
Baseline	Control	3427.82	1293.64	6115.13	1749.87	0.57	0.18	0.97	0.04	0.01	0.04
	Verbal	4224.00	1744.89	6755.35	2367.69	0.63	0.18	0.97	0.06	0.01	0.04
	Visual	4100.11	1315.79	6356.48	1627.51	0.65	0.16	0.98	0.05	0.01	0.03
Cue manipulation	Control	3494.46	1478.70	6019.83	2178.69	0.59	0.19	0.97	0.05	0.01	0.05
	Verbal	4391.80	1766.43	7990.51	2136.46	0.55	0.19	0.97	0.05	0.01	0.04
	Visual	7405.52	2199.37	12061.82	2920.55	0.62	0.15	0.97	0.04	0.01	0.06
All Trials	Total	4530.76	2145.61	7583.94	3069.63	0.60	0.18	0.97	0.05	0.01	0.05

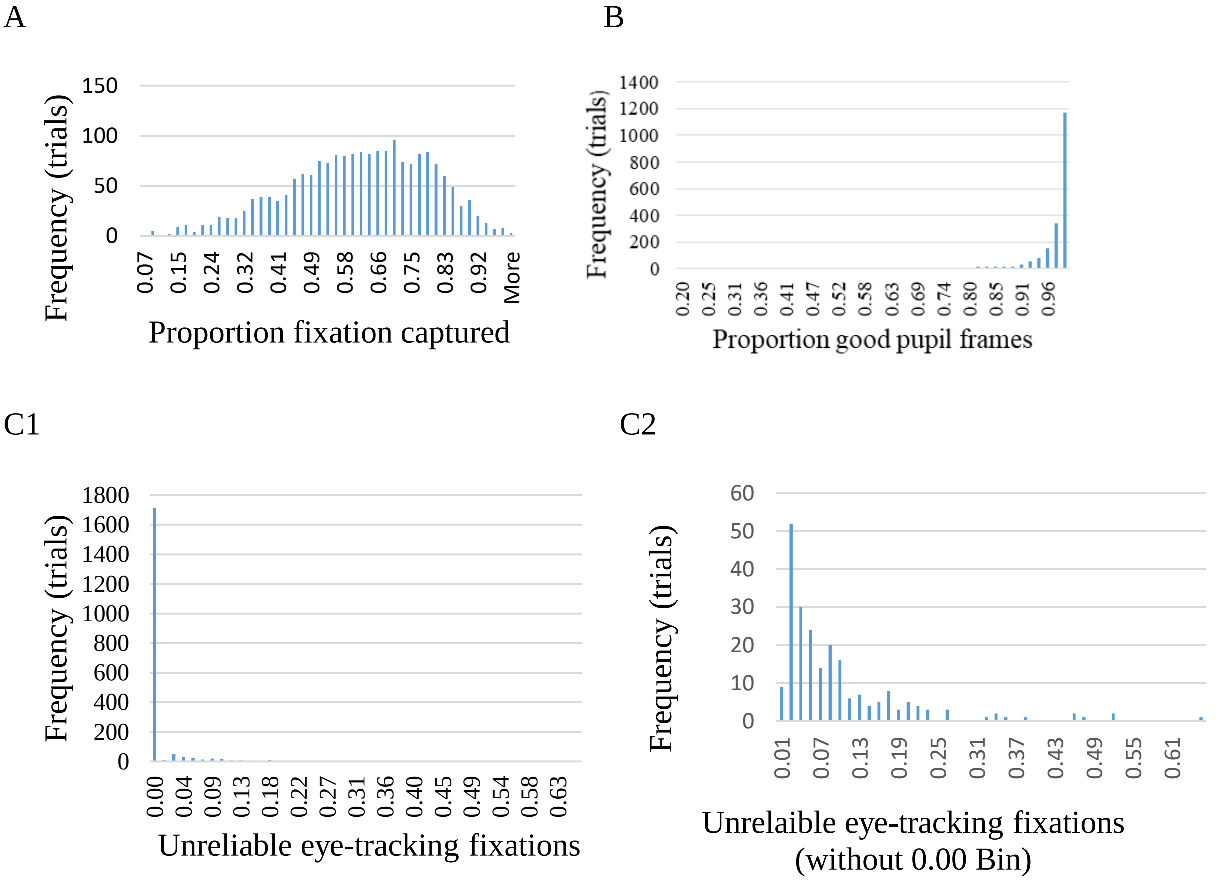


Figure B1. Histograms plotting frequency of trials with: A) percentage fixation captured (fixation duration/trial duration); B)proportion good pupil frames; C) and proportion unreliable eye-tracking fixations (subjectively coding). Panel C1 shows graph with all trials included, and Panel C2 shows graphs with only trials with unreliable eye-tracking fixations included.

Appendix C

Table C1

Effects of age, gender, block, and condition on probability correct for data without eye-tracking exclusions

Variable	χ^2	***df***	***p***	***Odds ratio***	***95% CI odds ratio***
Age	15.37	1	<.001	2.06	1.43, 2.99
Gender	0.02	1	.914	0.97	0.69, 1.37
Block	33.70	3	<.001		
linear	18.61	1	<.001	1.90	1.42, 2.54
quadradic	17.34	1	<.001	1.59	1.28, 1.97
cubic	0.27	1	.604	1.08	0.78, 1.57
Condition	44.81	3	<.001		

Note. Tukey posthoc test following up on the effect of condition showed the following differences in performance: Verbal > Visual, Control, Baseline; Visual >Baseline, Control = Baseline.

Table C2

Effects of age, gender, block, and condition on proportion fixations to target landmark using fixation duration

Variable	***b***	***SE***	***F***	***df***	***p***
Age	.05	.02	6.57	1,73.69	.012
Gender	-.01	.02	0.38	1, 73.9	.538
Block			1.58	3,1507	.192
Condition			473.33	3, 1349.61	<.001

Note. Tukey posthoc test following up on the effect of condition showed that Visual > Verbal > Control = Baseline

Table C3

Effects of age, gender, proportion landmark fixation per trial using trial duration and condition on probability of correct search

Variable	χ^2	***df***	***p***	***Odds ratio***	***95% CI odds ratio***
Age	13.02	1	<.001	1.87	1.33, 2.66
Gender	<.01	1	.957	0.99	0.72, 1.36
Prop. Landmark Fixation per Trial	7.08	1	.007	3.52	1.39, 9.02
Condition	33.22	3	<.001		
Prop. Landmark Fix. x Condition	4.80	3	.187		

Note. Tukey posthoc test following up on the effect of condition showed that Verbal > Visual = Control = Baseline).

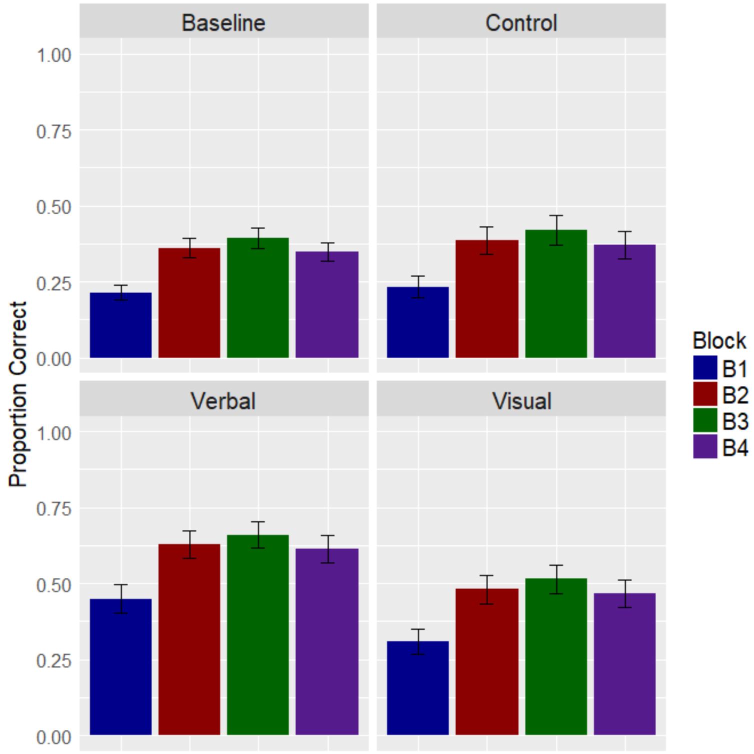


Figure C1. Proportion correct by condition and block for data without eye-tracking exclusions, controlling for effects of age and gender. Error bars are plotted +/- 1 standard error for point estimates from the logistic mixed-effects model (black lines). Model plotted with all data, including data with eye-tracking exclusions.

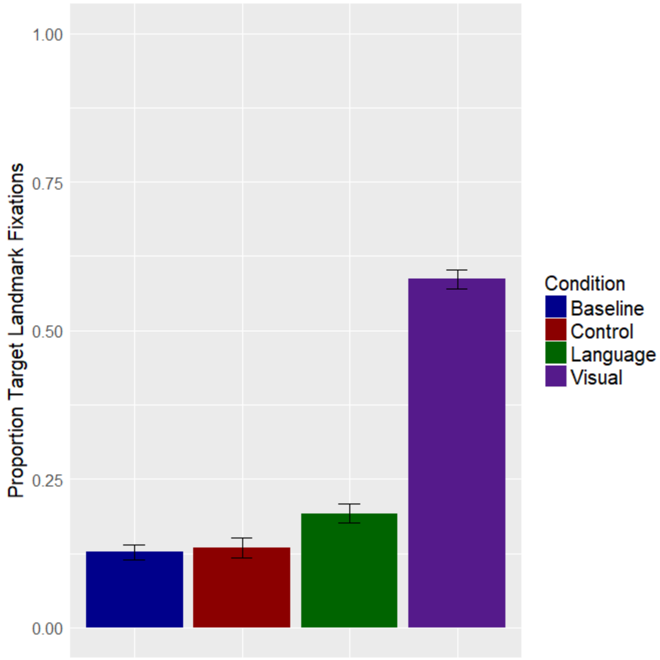


Figure C2. Proportion target landmark fixations by condition controlling for effects of age and gender. Proportion of target landmark fixations was calculated as the duration of fixation to target landmarks divided by the duration of on-task fixations. Error bars show +/- 1 standard error for point estimates from the mixed-effects model.

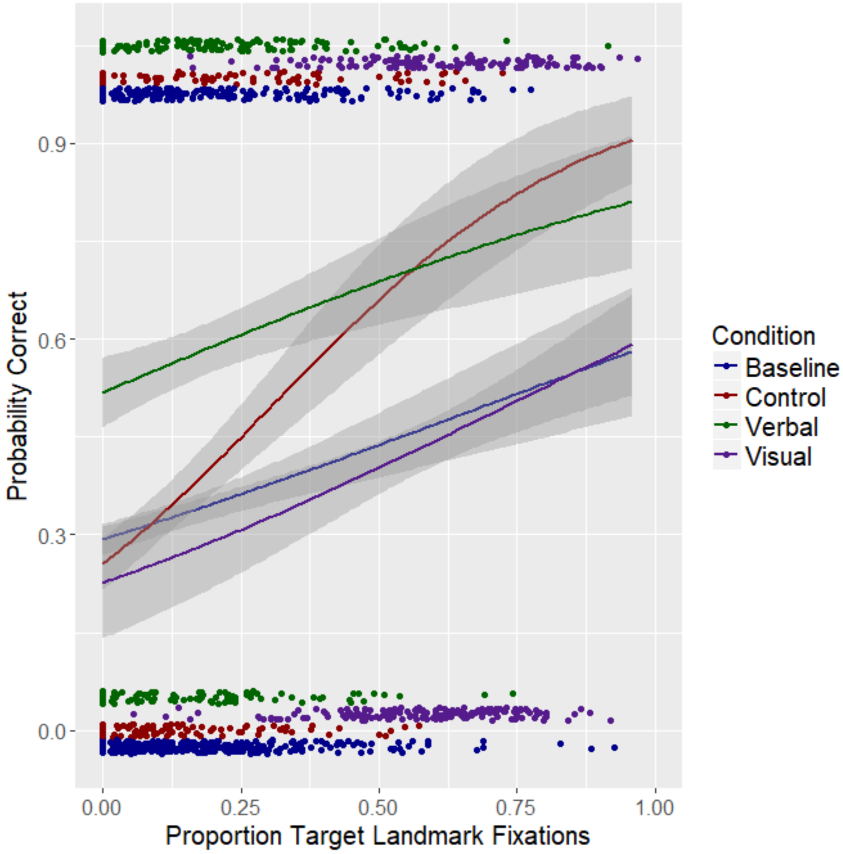


Figure C3. Probability correct by proportion target landmark fixations duration and condition, controlling for effects of age and gender. The data points represent individual trials for each condition indicating whether the trial was correct (1) or incorrect (0), with conditions offset vertically for visibility. Grey error bands are plotted +/- 1 standard error for point estimates from the logistic mixed-effects model.

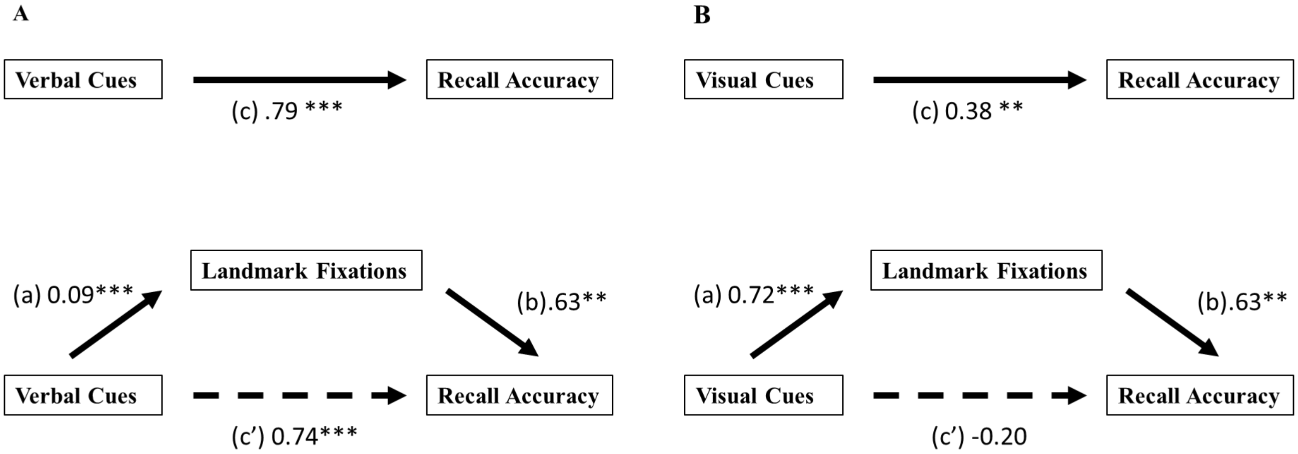


Figure C4. Standardized regression coefficients for the relation between verbal cues (A) visual cues (B) with baseline as the reference level and recall accuracy, mediated by proportion duration of target landmark fixations, controlling for effects of age and gender. The estimated indirect effect for the verbal condition was *b* = .02, 95% CI [<.01, .03] and the direct effect was *b* = .22, 95% CI [.11, .33]. The estimated indirect effect for the visual condition was *b* = .15, 95% CI [.05, .23] and the direct effect was *b* = -.02, 95% CI [-.13, .09].