Spring 4-22-2016

Preschoolers' Use of Verbal and Musical Strategies for Solving a Spatial Reasoning Task

Alyssa Marie Iulianetti
alyssa.iulianetti@student.shu.edu

Preschoolers' Use of Verbal and Musical Strategies for Solving a Spatial Reasoning Task
by
Alyssa Marie Iulianetti

B.A. Psychology, Immaculata University, 2014

A Thesis Submitted in Partial Fulfillment of the Requirements for the Master of Science in Experimental Psychology

In

The Department of Psychology
Seton Hall University
April, 2016

SETON HALL UNIVERSITY
College of Arts & Sciences

APPROVAL FOR SUCCESSFUL DEFENSE

Masters Candidate, Alyssa Iulianetti, has successfully defended and made the required modifications to the text of the master's thesis for the M.S. during this Spring Semester 2016.

THESIS COMMITTEE
(please sign and date beside your name)

Mentor:
Amy S. Joh: ______ 4/22/16

Committee Member:
Marianne Lloyd: ______ 4-21-16

Committee Member:
Paige Fisher: ______

Acknowledgements

I would like to thank my advisor, Dr. Joh, and the members of my committee, Dr. Lloyd and Dr. Fisher, for their guidance over the past two years. I would also like to thank the undergraduate research assistants who helped me to collect and code data, Alexis Finley, Anglin Thevarajah, and Stephanie Koeller.

Table of Contents

List of Figures

List of Tables

Abstract

Spatial reasoning is the ability to visualize and interact with one's three-dimensional environment. A spatial problem that is particularly challenging for young children is the tube task, developed by Hood in 1995. Children are presented with an apparatus that consists of a large frame with three intertwining tubes attached to the top and bottom. Children are asked to predict where a ball will emerge when it is dropped down one of the tubes. Young children typically make what is known as the gravity bias error, in which they believe a ball will drop straight down even though its path is obstructed by a curved tube. The current research consists of two studies with participants between 3 and 3.5 years of age. In both studies, the goal was to provide children with a strategy to help them overcome the gravity bias error. In Study 1, participants were asked to explain their predictions about the location of the ball at the start of each test trial. Children's responses in Study 1 were used to create the lyrics for the song in Study 2 in which participants were asked to sing a short song designed to facilitate reasoning about their predictions. Study 1 showed a pattern similar to previous research, with boys outperforming girls. It did not appear that the manipulation used in Study 1 resulted in improved performance. In contrast, Study 2 indicated that with this type of musical training, girls make the same number of correct predictions as boys, suggesting that this manipulation may be of use—albeit limited in scope—for girls. Together, the findings from the two studies provided further insight into how children reason about spatial problems and the types of external cues may help them solve these important problems.

Introduction

Spatial reasoning refers to the ability to perceive, visualize, and manipulate one's three-dimensional surroundings (Stiles-Davis, 1988). Spatial reasoning is a skill that is required for even the most mundane, everyday tasks. For example, several times a day, we use a fork to eat a meal. To eat with a fork, one must reach for and grasp the fork by orienting the hand in an appropriate way relative to the fork, move the fork in order to use it to pick up food, and then bring the utensil with the food to one's mouth in a manner that will result in getting the food into the mouth.

Spatial reasoning is a complex psychological ability involving various perceptual and cognitive skills. Halpern and LaMay (2000) identified five interrelated, measurable subcategories of visual-spatial abilities including spatial perception, mental rotation, spatial visualization, spatiotemporal ability, and generation and maintenance of a spatial image. Other researchers have identified broader categories of spatial skills, such as perspective taking and mental rotation (Newcombe & Frick, 2010). Regardless of the exact definition, researchers agree that skills such as visualizing and predicting the trajectory of moving objects also involve obtaining and using relevant perceptual information, problem-solving, and learning; they further agree that these skills develop gradually over the course of early childhood. The current research focused on how children think and talk about a challenging spatial problem, and whether explaining their predictions aloud (Study 1), or using a mnemonic-based musical strategy (Study 2) can help children solve this spatial problem.

Spatial Reasoning

Long before children begin to demonstrate observable spatial reasoning skills, they show hints of this complex ability with anticipatory eye gazes (Johnson, Amso, & Slemmer, 2003; von

Hofsten, 1980), mental rotation (Moore & Johnson, 2008; Moore & Johnson, 2011; Örnkloo & von Hofsten, 2007; Quinn & Liben, 2008), and spatiotemporal reasoning (Kim & Spelke, 1999). Spatial skills continue to develop gradually and strengthen throughout early childhood (for a review, see Newcombe & Frick, 2010). However, one spatial task that remains challenging for young children is Hood's (1995) tube task, which is used to examine children's spatial reasoning abilities with three intertwining opaque tubes (Figure 1). In this task, a ball is dropped down one of three "chimneys" connected to the tubes, and children are asked to search for the location of the ball at the end of one of the three tubes. Several studies have shown that children under the age of four consistently choose the opening directly below where the ball was dropped (point C in Figure 1) (Hood, 1995; Hood, 1998; Hood et al., 2000; Freeman et al., 2004; Hood et al., 2006; Bascandziev & Harris, 2010, 2011; Joh et al., 2011; Joh & Spivey, 2012). Children learn early on that the typical trajectory for a falling object is straight down if there are no external obstructions (Hood, Carey, & Prasada, 2000). However, in the tube task, the curved tube obstructs this vertical path. Nevertheless, young children are generally unable to reason about this change to the trajectory of the ball. Hood (1995) conjectured that this response is driven by a "gravity bias" because children under four years of age expect an object to fall in a vertical line even when it is impossible for the object to do so.

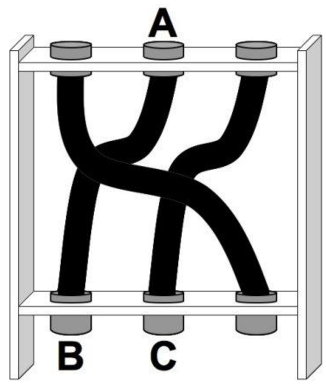

Figure 1. Apparatus for tube task, which contains tubes arranged in a non-vertical fashion. For example, when a ball is dropped through point A, point B is the correct choice. However, children under the age of four typically search for the ball at point C (gravity bias error).

Hood (1998) then attempted to determine whether children's inability to solve the tube task was due to a gravity bias or the proximity of the opening to the spot where the ball was dropped. Using video, he reversed the direction of the ball so that gravity was no longer a factor. Children no longer made the vertical line error when the ball started at the bottom of the tube, confirming that their errors were due to a gravity bias. Hood, Santos, and Fieselman (2000) then determined that while many errors were still made, children were more likely to succeed at the tube task when the ball was rolled horizontally rather than vertically. This also suggests that vertical spatiotemporal reasoning is a challenge for young children because of the gravity bias.

Since Hood's original studies, further research has been performed to determine the factors related to children's success on this spatial task. Thus far, researchers have determined that the development of inhibitory control as well as additional visual or verbal information can help children overcome the gravity bias error. Additionally, sex appears to matter: boys tend to outperform girls on this task. (Hood, 1995; Joh & Spivey, 2012).

Inhibitory control. A primary issue that has been theorized to contribute to the gravity bias error is inhibitory control, which is the ability to inhibit responses or ideas that are not relevant to the task at hand (Carlson, Moses, & Claxton, 2004). Within the context of the tube task, young children cannot seem to suppress their initial belief that objects will fall downward vertically because of gravity.

One way to examine the progression of inhibitory control is to ask children who no longer make gravity bias errors to *avoid* where the ball would land by predicting where the ball would *not* emerge. Freeman, Hood, and Meehan (2004) chose this sample to examine whether inhibitory control was still a factor in solving the tube task even when they were able to find the correct answer. With this version of the tube task, there are two correct answers: the gravity bias location and the third (miscellaneous) location. Four-year-old children actively avoided the gravity bias location, indicating that the inhibitory control to avoid this opening is still prevalent children at this age. Only by 7 years of age were children able to fully understand the task and therefore showed no preference for either of the two openings of which the ball would not come out. This full control over inhibition allowed for a level of certainty in the two incorrect openings that was not displayed by younger children.

Hood, Wilson, and Dyson (2006) further found that 4-year-old children, who are generally able to overcome the gravity bias, were more likely to make the gravity bias error when two balls were dropped down two of the tubes simultaneously. It appeared that this return to gravity bias responses was not due to an inability to track two balls, but rather taxed inhibitory control. If children were unable to track two balls, they would be making random errors as opposed to remembering where the ball was dropped and consistently choosing the opening directly underneath. With the second ball, 4-year-old children were unable to inhibit their initial gravity-driven response: the opening directly below where the ball was dropped.

Another method of examining inhibitory control is to measure conflict inhibition, or selecting between two conflicting rules in a task. Baker, Gjersoe, Sibielska-Woch, Leslie, and Hood (2011) found that conflict inhibition accounted for a significant amount of the variance in 3-year-olds' performance on the tube task. In this study, children were asked to manually search in one of the three openings for the ball. Thus, the researchers postulate that this type of manual search requires similar skills to those used in conflict inhibition. The tube task presents two conflicting solutions: the gravity bias and the correct solution. The better children are at inhibiting their initial response to look for the ball directly under where it was released, the greater their skill in conflict inhibition. This research provides further evidence that inhibiting prepotent responses enhances performance on the tube task.

Similarly, Spivey and Joh (in preparation) found that children's ability to make correct predictions on the tube task was moderated by how quickly they could learn inhibitory rules. This was measured by 3 different types of inhibitory control measures: the day-night task, which is an age-appropriate version of the Stroop task, the card sort task, which asks children to sort cards based on alternating rules, and the frog-monkey task, in which children must follow the instructions given to them by one puppet but not another. Taken together, these studies provide evidence that inhibitory control is an important development that contributes to success on the tube task.

Visual information. In addition to inhibitory control, the availability of visual information appears to influence children's success in the tube task. For example, Bascandziev and

Harris (2011) attempted to determine whether the "chimneys" on top of the tubes were hindering children's ability to predict the correct location of the ball. Children who were first provided the task with the inclusion of chimneys showed improved performance when the chimneys were removed. Conversely, children who were initially provided no chimneys showed decreased performance when chimneys were later attached. This finding indicates that the chimneys may have provided a confusing visual cue, and thus children improved when they were able to see the ball being dropped directly into the tube.

Joh and Spivey (2012) provided a different type of visual information: color cues. The researchers found that the presence of color cues as visual information also appears to help children make correct predictions in the tube task. Children were tested with three tubes that were each a distinctly different color. The intent of the color cues was to help children obtain a strategy with which to inhibit the gravity bias. The colors would make it easier for children to trace the path of the tube in order to make a correct prediction. However, while the colored tubes did help them choose the correct opening, it did not teach them how to solve the task correctly once the cues were removed. Therefore, children were not using the cues to overcome the gravity bias, but rather used the cues to solve the problem for them (e.g., children were simply matching the color at the top to the color at the bottom without considering the path of the tube). Generally, visual information may provide helpful cues for solving the tube task, but may not actually teach children about how to solve the task without the presence of these cues.

Verbal information. Finally, hearing someone else talk about the tube task in a manner which draws attention to the path of the tube may help children overcome the gravity bias. Bascandziev and Harris (2010) tested whether 2.5- to 3.5-year-old children could make correct predictions by providing a verbal testimony that drew their attention to the top and bottom of the tubes. They compared this against merely drawing attention to the tubes but giving no information about the path of the ball. They found that 3.5-year-old children were generally able to overcome the gravity bias with a verbal testimony about the top and bottom of the tube. In a second experiment, 3.5-year-old children were divided into three conditions that involved verbal testimony: no escape, which described the fact that the ball cannot escape the tube ("The ball could not escape from that tube. It rolled inside the tube."), eye movement, which instructed children to follow the tube with their eyes ("What you need to do is to watch which tube the ball goes in and then you need to follow that tube with your eyes. Okay?"), and attention, which gave non-specific information about the ball traveling down the tube ("You have to pay attention to the tubes in order to find the ball immediately."). The results show that the type of verbal instruction is critical, as both the no escape and eye movement conditions improved performance, but the nonspecific attention condition did not.

Similarly, Joh, Jaswal, and Keen (2011) compared children's performance when providing specific verbal information to nonspecific verbal information. Specifically, they found that the word "imagine" seems to trigger a change in the way in which children think about the tube task. Children were able to overcome the gravity bias significantly more frequently when asked to "imagine the ball rolling down the tube" than when given a different verbal testimony without the word "imagine," known as the "wait" condition, ("The ball is going to roll down the bumpy tube") or a control in which no testimony was given. The "wait" condition was included to ensure that it wasn't simply the extra time provided by the experimenter talking that improved performance. This finding indicates that it is possible that children benefit from the prompt to visualize the problem in order to overcome the gravity bias. Additionally, the fact that the verbal condition without the word "imagine" ("wait") yielded no improvement over the control condition indicates that the success of the “imagine” condition is not related to the extra time to think provided by the verbal instruction. Furthermore, children in the control and “wait” conditions rarely switched their choices. However, children in the “imagine” condition did frequently switch (i.e., held the ball under one opening, then changed their answer to a different opening before the ball was dropped), as if the instructions prompted children to second-guess their initial decisions. This indicates that after children made their initial prediction, which was generally a gravity bias error, the “imagine” instruction encouraged them to rethink their predictions and increased the number of times they switched to a different answer. While it is still unclear exactly what encourages children to make correct predictions, it is clear that specific types of verbal cues that focus on the path of the ball rolling through the tube help them greatly with this task.

Sex differences. Findings of sex differences on the tube task are inconsistent. Hood (1995) found a significant main effect on sex for the tube task: namely, boys outperformed girls on this task. This aligns with previous findings that boys generally perform better on spatial reasoning tasks than girls (for a review, see Halpern & LaMay, 2000). Joh & Spivey (2012) did find a marginal effect of sex in that boys made more correct predictions than girls on the tube task. Furthermore, there was a significant main effect on switching: boys switched the placement of the cup before the ball was dropped more frequently than girls. This indicates that boys were more likely to second guess their typically incorrect initial responses and switch to a correct answer. However, Bascandziev & Harris (2010; 2011) found no main effect of sex. A more comprehensive review of tube task studies, varying in type (motor, visual, verbal) and amount (0%, 50%, 100% of trials) of external information provided to 273 participants, showed that not only did boys significantly outperform girls, but also benefitted from several types of training more than girls (Joh, under review). These findings suggest that, like most other spatial tasks, boys are likely to outperform girls.

Verbalization

Apart from the verbal testimony of others, and in a broader context than that of spatial reasoning, another way in which researchers have attempted to examine how children solve problems is through children's own verbalizations during difficult tasks. These verbalizations can take the form of private speech, which is an externalization of a child's inner thoughts, or social speech, which is speech directed toward another person.

Private speech. It has been reported that children as young as 2 years of age use spontaneous nonsocial verbalization, otherwise known as private speech, to facilitate problem solving (Manning, White, & Daugherty, 1994). Indeed, nonsocial talking about a problem seems to be a prominent feature in young children. According to Gredler (2009), Vygotsky theorized that private speech develops in three stages between the ages of 3 and 7. The first stage consists of emotional speech that often appeals to the object of the problem (e.g., "The blocks should stop falling"). The second stage involves comments that accompany a child's actions ("I am stacking the blocks"). The third stage implements verbal planning of actions ("I am going to put the big blocks on the bottom"). Private speech peaks at 4 to 5 years of age, and then gradually decreases until it becomes internalized into inner speech. Private speech is generally fully internalized by the time children are 7 to 8 years of age. (Manning, White, & Daugherty, 1994).

Private speech is believed to be a sign of cognitive processing because it varies with the difficulty of a task. Verbalization increases with difficulty, but once a task becomes too difficult the verbalization levels off or decreases. At the optimal difficulty level, that is, the task is very challenging for the child but not impossible, the most verbalizations are seen (Gever & Weisberg, 1970). Additionally, private speech is more likely to occur at the beginning of a task. Goodman (1984) studied 3.5- to 5-year-old children and coded their use of speech during a puzzle task. Private speech occurred most frequently at the beginning of the task, and was associated with the difficulty of the task. That is, the more difficult the task, the more likely a child is to talk to him or herself about the problem in an effort to solve it.

Social speech. While young children frequently use private speech to work through problems, they may also be able to answer questions about their reasoning during problem solving. Social speech is verbalization directed toward others instead of oneself. Three- and four-year-old children are capable of verbalizing their reasoning for decisions made during a spatial task that involves taking a perspective other than their own. Ives (1980) asked children to describe different perspectives of an object either verbally or using pictures. Children who verbalized their reasoning were more likely to correctly identify a perspective other than their own. Through both private and social speech, verbalization appears to increase children's ability to solve challenging problems.

A major development in the progression of social speech for preschool-aged children takes place through pretend play, which gives children a context through which they can work through and reason about complex situations. Buchsbaum, Bridgers, Weisberg, and Gopnik (2012) presented 3- and 4-year-olds with a scenario in which some toys could make a machine play music (labeled as "zandos"), while others could not ("non-zandos"). The participants were then asked counterfactual questions about the scenario ("If this one were not a zando, what would happen if we put it on top of the machine?"). When asked to use pretend play to work through the scenario counterfactually ("Which of these should we try to pretend to make the machine play music?"), children were better able to reason correctly about these counterfactual questions. Furthermore, their pretend play aligned with the causal rules set up by the novel relationship introduced, providing evidence of a strong relationship between the ability to play and the ability to reason counterfactually. It may be that children benefited from the pretend play context which encouraged them to work through the problem via social speech.

How children benefit from verbalization. Private and social speech may help children solve problems by focusing their attention on relevant aspects of the problem. For example, children can be taught spatial language such as "left" and "right," as well as how these words relate not only to their own bodies but also to their environment. In one study, 4-year-olds were taught similar spatial words then asked to solve a reorientation task that children are otherwise typically incapable of solving at that age: using spatial language cues to find a hidden object. Using a red wall as a landmark, children would be given clues such as "The toy is to the left of the red wall." Children at this age were able to use spatial language cues to solve the task (Shusterman & Spelke, 2005). Additionally, when children are asked to use spatial language to describe a path or action, they are more likely to include descriptions of the path that look toward the goal and neglect to describe the source and its path (Lakusta & Landau, 2005). For example, if children were asked to describe themselves walking from their house to the mailbox, they would more likely say "I am walking to the mailbox" rather than "I am walking away from the house." This type of goal-directed spatial language indicates that from a young age children are able to verbally focus on the relevant aspects of a spatial task.

Focusing children's attention to relevant components of a problem may help them remember solutions and strategies. For example, DeLoache, Cassidy, and Brown (1985) examined emerging mnemonic strategies in 18- to 24-month-old children. Children played a hide and seek game in which they were shown the hiding spot of a toy and told they would have to retrieve it later. Some of the target behaviors that showed early emergences of mnemonic strategies included verbalizing, looking, pointing, approaching the toy, peeking at the toy, and attempting to retrieve the toy early. A mnemonic device is a strategy used to help improve memory. Mnemonics are most commonly described as devices used for rote memorization such as “ROYGBIV” to remember the colors of the rainbow. Much like other verbalizations, the usage of these mnemonic strategies was related to the perceived difficulty of the task. That is, the more difficult the task was thought to be, the more mnemonic strategies emerged. Researchers found that when multiple toys needed to be retrieved, objects which were the focus of mnemonic strategies were more likely to be successfully retrieved. Thus, verbal and motor mnemonic strategies improved performance on a problem solving task.

Music and Mnemonics

One technique for learning and problem solving that has not yet been applied to spatial problems is the incorporation of a musical mnemonic device. Music is a prevalent part of children’s everyday lives. Music is included in children’s television programs, electronic games and apps, group activities at school or day care, and tools for learning, such as remembering the alphabet or how to complete certain tasks. Children spontaneously include music in their everyday lives because it is a fun, accessible way for them to interact with and learn about the world around them. For these reasons, music may be an effective mnemonic tool to provide verbalizations that will help young children acquire problem solving strategies for difficult tasks.

Rainey and Larsen (2002) delineate three criteria for a successful mnemonic device: it must create structure for the information, it must be easy to recall, and it must facilitate the retrieval of the information. Similarly, Moore et al. (2008) suggest that information is more easily remembered and manipulated when presented in an organized manner. Thus, mnemonic devices, including musical mnemonic devices, should allow for information to be easily remembered. Music facilitates rote memorization, and thus can be used as a mnemonic strategy, but it is hypothesized that music goes beyond this rote memorization to give children a context through which they can talk through difficult problems, as well as recall strategies for problem solving, similar to that of pretend play.

Effects on cognitive development. Research has found that musical training can improve spatial-temporal reasoning over time (Habe & Jausovec, 2003; Rauscher, Shaw. & Ky, 1995; Rauscher et al., 1997; Raunchier & Zupan, 2000). However, little research has been conducted on music's immediate effect on cognitive development in the short-term. Research has shown that adults have improved recall ability when music is used as a mnemonic strategy (Rainey & Larsen, 2002; Thaut, Peterson, McIntosh, & Hoemberg, 2014). However, less research has been performed examining the short-term effects of music for young children. But promisingly, Gfeller (1983) found that when enough repetition was provided, both typically developing and learning disabled 9- to 11-year-old children performed significantly better on a memory task when provided with a musical mnemonic strategy rather than a verbal mnemonic strategy. Although the research has indicated that musical strategies require more time to learn, the memory benefits are greater once the strategy is learned.

Another study determined that early music training is correlated to visual memory and abstract thinking skills. Bilhartz, Bruhn, and Olson (1999) examined the effects of music training on the cognitive development of 4- to 6-year-olds. They found that early music training was significantly correlated to the Bead Memory subtest of the Stanford-Binet Intelligence Scale when controlling for sex, ethnicity, parents' education level, and economic status. The Bead Memory test involves visual memory and abstract thinking, and is not music-related at all. In this test, participants are required to remember and utilize beads of different colors and shapes. Because the test is unrelated to music training, the correlation suggests a link between the cognitive processes used in music training and this memory test.

Music encourages creativity in problem solving, especially for abstract concepts (Hitz, 1987). Exploring music allows children to ask questions and think about tangential concepts. Much like private speech, music provides children with a tool for externalizing their inner thoughts to help them work through a problem. Furthermore, rhythm enhances attention, and in turn learning, in educational settings (Geist & Geist, 2012).

However, Calvert and Billingsley (1998) examined preschoolers' ability to recall and comprehend information presented in a song versus a verbal context. They found that with repetition, children were able to accurately recite information better when presented as a song, even if this song was not in their native language (English). However, age predicted comprehension for the meaning behind these songs. These findings indicate that younger children, while still able to memorize the information, use a shallower level of information processing than older children.

Calvert (2001) further examined the impact of songs in educational television programs and found that music helped improve rote memorization, but a deeper level of understanding was associated with verbal, non-musical instruction. This finding suggests a need to combine music and language in order to maximize comprehension and memory. Calvert suggested that music can be encoded at a shallower level of processing, and that songs may improve rote memorization but may not necessarily facilitate deeper level, semantic processing. The implications of the study indicates that repetitive songs can improve some types of rote memory and potentially visual sequencing tasks if the songs are in a television program. However, spoken information was more effective for the internalization of the factual information provided. Educational songs may not necessarily facilitate deeper levels of processing of the material.

Current Studies

Two studies explored spatial reasoning in preschool-age children, focusing on whether verbal or musical mnemonic strategies could improve their performance, similar to the verbal strategies used by Bascandziev and Harris (2010) and Joh et al. (2011). In Study 1, children were asked, “Why do you think the ball will come out of there?” as they participated in the tube task. The question was posed after they made their initial prediction, but before the ball was dropped, to encourage children to verbalize their solutions. In Study 2, children were encouraged to use a song as a mnemonic strategy during each trial before they were asked to make a prediction. The song was created from the key words children provided in Study 1. In both studies, an adaptation of Hood’s tube task (similar to the one used by Joh et al., 2011, and Joh & Spivey, 2012) was used because it is a developmentally-appropriate spatial task that is challenging for young children. The difficulty of the task should lead to various problem-solving attempts, including verbalizations. Moreover, the difficulty of the task indicates that if the verbal or musical strategies are helpful for children, then children should show measurable improvements in performance. Participants also completed the Expressive Vocabulary Test, Second Edition (EVT-2), in which expressive vocabulary levels were measured by asking children to label what they see in a flip book of familiar pictures. Because both tasks relied heavily on verbalization-based responses, children's EVT scores were used to provide a standardized score to compare their verbal capabilities to their peers.

In both studies, participants were 3- to 3.5-year-old children, since at this age they generally have not yet overcome the gravity bias but are old enough to express and understand verbal and musical instruction (Bascandziev & Harris, 2010; Joh et al., 2011; Geist & Geist, 2012). Two-year-old children are the most susceptible to the gravity bias, however, their language skills are not as developed and therefore they may not fully understand the instructions given. By four years of age, children are beginning to overcome the gravity bias. Three-year-old children are still likely to make the gravity bias error, and are more likely to understand language-based instruction. Furthermore, previous research with the tube task has indicated that children at this age will improve with external help such as verbal information (Bascandziev & Harris, 2010; Joh et al., 2011).

The primary measures of interest were children's correct predictions and their rates of switching. Correct predictions, not gravity bias errors, were measured because correct predictions are a more stringent measure of performance—they include errors due to gravity bias and miscellaneous factors. In previous studies, help in the form of visual or verbal cues led to increased correct predictions (Bascandziev & Harris, 2010; Bascandziev & Harris, 2011; Joh & Spivey, 2012). Switching indicates a measure of hesitation or 'thinking through the problem' by trying out alternative choices. In previous studies, conditions which provided children with opportunities to think about their predictions led to increased switching (Joh et al., 2011).

It was predicted that in Study 1, children who were able to verbally rationalize their predictions with appropriate key words would make more correct predictions. It is also possible that hearing the "why" question would encourage children to switch their predictions during the course of the trial. Regardless of the outcome of Study 1, the verbalizations would help create a picture of how children at this age think about the task, and thus help to shape the lyrics for Study 2. In Study 2, it was predicted that children who were able to use the mnemonic song, that is, children who sang along to at least part of the song, would make more correct predictions. It is possible that the song could lead to increased or decreased rates of switching. If the song makes children question the problem, like the "why" question in Study 1, then children in Study 2 should also show increased rates of switching. However, if the song provides children with answers without encouraging them to second-guess the task, then children may show lower rates of switching in Study 2. Unlike Study 1, Study 2 may not make children second-guess their initial decisions, but rather help them make more informed guesses before they make their first prediction. It was also predicted in both studies that participants would improve over time, which would be measured by comparing the first half of the trials (trial block 1) to the second half of the trials (trial block 2).

Study 1: Verbalization and Spatial Reasoning

Method

Participants. Thirteen children between 3 and 3.5 years of age were recruited for Study 1 (8 girls and 5 boys, *M* age = 37.79 months, *SD* = 1.80). One additional participant did not complete the session due to fussiness, and therefore was excluded from analyses. Participants were primarily from middle class families and they were Caucasian (*n* = 10), African American (*n* = 1), and biracial (*n* = 2). Participants were recruited through flyers, websites, mail, community organizations, and word of mouth from the communities surrounding Seton Hall University. Children received a small gift (e.g., a ball) and a photo souvenir for their participation.

Materials.

Tube apparatus. To create the tube apparatus, a wooden frame measuring 62.5 cm high, 59.1 cm wide, and 8.9 cm deep was used for the sessions (Figure 1). This frame contains three openings at the top and three openings at the bottom. Each opening has a round, white, plastic "chimney" that measures 5.7 cm in diameter and 6.4 cm high. These chimneys allow the tubes to be securely connected to the frame. Three opaque, flexible plastic tubes (4.4 cm in diameter and 67.6 cm long) can attach from one of the top chimneys to one of those at the bottom to create a path for a small ball. The ball is made of hard foam, and is approximately 2 cm in diameter. A small, transparent plastic cup was used to catch the ball.

EVT-2. The Expressive Vocabulary Test, Second Edition (EVT-2) is a standardized test that was administered to determine participants' expressive language abilities. The participants were shown pictures from a booklet. The pictures were of familiar objects or events such as a dog or a girl singing. Participants' responses were recorded in one of two standardized forms provided in the EVT-2 kit (Williams, 2007).

The EVT-2 manual provides charts with which raw scores (number correct) can be converted into standardized scores based on age. Thus participants with varying ages can be standardized to the same measure, as well as compared to age norms as presented in the manual. Both reliability and validity have been assessed for this measure. Split-half reliability was measured with a mean of .94 for Form A and .93 for Form B. Alternate form reliability was also measured to have a mean of .87 across age groups. Test-retest reliability yielded a mean of .95. The EVT-2 has a mean correlation of .82 to the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), which measures receptive vocabulary and is meant to complement the EVT-2. Furthermore, correlation between this test and the original EVT yielded an average of .81 (Williams, 2007).

Recording. Behavioral and verbal responses were recorded using a Canon Vixia HF R52 HD flash memory camcorder and a Sony DCR-SR85 camcorder external microphone.

Procedure. Parents remained in the testing room with their children during the entire session. Prior to the start of the session, participants spent about 10-15 minutes playing with toys and games in the laboratory to acclimate to the testing environment.

Familiarization. Participants were familiarized with the tube, ball, and cup with which they can catch the ball separately. This is outlined in Figure 2 below. The experimenter showed the participants that the tube is hollow, and a ball can roll through it, by saying, “Do you see this tube? It’s empty inside. Because it’s empty inside, I can roll a ball through it. Can you catch the ball?” After the participants caught the ball, the experimenter asked the participants to roll it back through the tube (Tube Introduction and Practice, Figure 2).

The experimenter then demonstrated how the cup can be used to catch the ball, and how the participants could indicate that they are ready for the ball to be dropped. The experimenter said, “Do you see this cup? It’s also empty inside. We can use it to catch the ball. See? I got it!

Do you want to try? Tell me when you're ready for me to drop the ball!" The experimenter dropped the ball only after the participants had said, "Ready!" (Cup Introduction and Practice, Figure 2).

Participants were then introduced to the wooden frame before the tubes were attached and the experimenter first demonstrated the function of the "chimneys." The experimenter then showed how the tube, ball, and cup function in conjunction with the frame, saying, "Do you see these three chimneys up here? And do you see these three chimneys down here? They go with the ball and the cup!" The experimenter showed how the ball can be dropped through the chimney to be caught in the cup. The experimenter said, "If I hold the ball here, then I'll put the cup where I think the ball will come out, and I got it! Do you want to try?" (Chimney Introduction and Practice, Figure 2). Finally, participants were given the cup, and the experimenter held the ball above one of the openings and said, "I'm going to hold the ball here. Can you put the cup where you think the ball is going to come out and tell me when you're ready?" The ball was dropped after the participants said "Ready!" After the participants practiced this three times, once under each of the three openings, the experimenter reintroduced the tube and attached it to the chimneys.

Participants then practiced catching the ball with only one tube in the frame at a time, for all three positions of the tube (top left to bottom right, top middle to bottom left, and top right to bottom middle). The experimenter first demonstrated by holding the ball above the chimney with the tube attached, putting the cup under the bottom chimney where the tube is attached, and saying, "If I hold the ball here, I can put the cup where I think the ball is going to come out, and I got it! Do you want to try?" Then the experimenter said, "I'm going to hold the ball here. Can you put the cup where you think the ball is going to come out and tell me when you're ready?"

The ball was dropped after the participants said “Ready!” If the participants made an incorrect prediction on any single tube trial, the experimenter repeated the practice for that single tube configuration, up to three times per configuration. Incorrect predictions on the single tube trials were rare, occurring for only three participants (Single Tube Introduction and Practice, Figure 2).

Test trials. The experimenter inserted all three tubes into the frame immediately following the warm-up phase. The experimenter held the ball above one of the three top chimneys and asked the children to place the cup under the bottom chimney where they thought the ball would come out. As with the practice trials, the experimenter said, “I’m going to hold the ball here. Can you put the cup where you think the ball is going to come out and tell me when you’re ready?” After indicating that they were ready, the participants were asked, “Why do you think the ball will come out there?” After the participants had been given time to explain their reasoning, the experimenter would then drop the ball. The experimenter did not give any feedback or indications regarding whether the participants were correct in his or her verbalizations or predictions. This procedure repeated for twelve trials. The experimenter rotated the frame 180° and switched the opening through which the ball was dropped on each trial. This procedure is outlined in Figure 2 below.

Familiarization: Tube Introduction	Familiarization: Tube Practice	Familiarization: Cup Introduction	Familiarization: Cup Practice
The experimenter holds up a tube, saying, “Do you see this tube? It’s empty inside. Because it’s empty inside, I can roll a ball through it. Can you catch the ball?” The experimenter waits for the participant to hold out their hands, then rolls the ball through the tube.	The experimenter asks the participant, “Can you roll the ball back through the tube?” The experimenter catches the ball.	The experimenter holds up the cup, saying, “Do you see this cup? It’s also empty inside. We can use it to catch the ball.” The experimenter drops the ball into the cup and says, “See? I got it! Do you want to try?”	The experimenter holds the ball above the cup and says, “Tell me when you’re ready for me to drop the ball!” The experimenter drops the ball after the participant says, “Ready!”
Familiarization: Chimney Introduction	Familiarization: Chimney Introduction	Familiarization: Chimney Practice	Familiarization: Chimney Practice
The experimenter shows participants the wooden frame and introduces the chimney openings, pointing to each one and saying, “Do you see these three chimneys up here? And do you see these three chimneys down here? They go with the ball and the cup!”	The experimenter then drops the ball through one of the bottom chimneys and catches it in the cup, saying, “If I hold the ball here, I can put the cup where I think the ball is going to come out, and I got it! Do you want to try?”	The experimenter holds the ball above one of the bottom chimneys and says, “I’m going to hold the ball here. Can you put the cup where you think the ball is going to come out and tell me when you’re ready?”	After the participant has chosen an opening and said, “Ready!” the experimenter drops the ball. This repeats for all three bottom chimneys.
Familiarization: Single Tube Introduction	Familiarization: Single Tube Introduction	Familiarization: Single Tube Practice	Familiarization: Single Tube Practice
The experimenter attaches one tube to the top left chimney and bottom right chimney, holds the ball above the chimney where the tube is attached.	The experimenter then drops the ball through the chimney with the tube attached and catches it in the cup, saying, “If I hold the ball here, I can put the cup where I think the ball is going to come out, and I got it! Do you want to try?”	The experimenter holds the ball above the chimney where the tube is attached and says, “I’m going to hold the ball here. Can you put the cup where you think the ball is going to come out and tell me when you’re ready?”	After the participant says, “Ready!” the experimenter drops the ball. This repeats for all three tube configurations (top left to bottom right, top middle to bottom left, and top right to bottom middle).

Figure 2. Procedure for the test trials of Study 1.

EVT-2. Participants were then administered the EVT-2. Following the protocol in the EVT-2 manual, the experimenter asked questions such as “What do you see?” with questions and pictures increasing gradually in difficulty (e.g., labeling only part of the picture, describing an action, or providing a synonym). Participants were asked to provide one word responses. As per the protocol, the experimenter said “Very good!” after every answer, regardless of whether or not the participants provided a correct response. The test booklet includes a labeled list of words for each trial which, if given by the participants, requires the experimenter to prompt for a more specific answer. The experimenter recorded the participants’ responses in a test booklet. The booklet lists what counts as a correct or incorrect response, and whether a prompt for a more specific answer is required. Testing continued until the ceiling level (5 incorrect responses in a row) was reached. From this, a standard score was calculated and used as a measure of expressive vocabulary ability. Two forms, A and B, were counterbalanced between participants (Williams, 2007).

Data coding. Each session was coded in Excel for participants’ *first prediction* (the first opening under which the cup was held for two or more seconds) and *final prediction* (the opening under which the cup was held for two or more seconds when the ball was dropped). The first prediction could be the same or different from the final prediction, depending on whether participants *switched* (changed their answer and held their cup underneath a different opening for more than two seconds). If no switch occurred, then the initial and final prediction was the same. Switches were only analyzed if the initial and final predictions were different. A switch could take place at any point during the trial, before or after participants were asked why they chose a certain opening. Whether a switch was made before or after the manipulation was also coded. Experimenters also transcribed what the participant said verbatim in response to the question, “Why do you think it will come out there?”

The primary coder scored all trials for all participants using the codes described above. A reliability coder scored all trials for 33% of participants using the same codes. The two coders agreed on 92% of trials for first predictions, 98% of trials for final predictions, and 100% of trials for number of switches. Disagreements were resolved through discussion.

Study design. The three measures of interest were the number of initial correct predictions, final correct predictions, and switches. Trial block (trials 1-6 vs. trials 7-12 and thus, a within-subjects variable) was examined to determine whether participants improve over time. Sex (male vs. female and thus, a between-subjects variable) was analyzed to examine if there is a sex difference in performance on the tube task. Because participants' ages ranged from 36.06 to 41.75 months, age was controlled as a covariate. Additionally, because participants' EVT scores ranged from 108 to 134 (70^{th} to 99^{th} percentile for this age group), verbal ability was controlled as a covariate.

Results and Discussion

As in previous studies, preschool-aged children made mostly gravity bias errors. On their initial predictions, participants made correct predictions on 28.2% of trials, gravity bias-driven predictions on 63.5% of trials, and miscellaneous (third opening) errors on 8.3% of trials. Although participants showed improvements between initial and final corrections, the patterns remained similar for final predictions. Participants made correct final predictions on 37.2% of trials, gravity bias errors on 55.8% of trials, and miscellaneous errors on 7% of trials. In general, these responses are comparable to those from control conditions in previous studies (Joh et al., 2011; Joh & Spivey, 2012), suggesting that asking the "why" question did not significantly improve spatial performance.

To further examine which variables, if any, influenced children's performance, a 2 (trial block) x 2 (sex) mixed-design ANCOVA was conducted separately for initial and final correct predictions. Test age and EVT scores were entered as covariates. Neither ANCOVA revealed significant main effects or interactions, suggesting that the "why" question manipulation did not result in improved performance across trials blocks or that EVT scores or test age were related to performance (all *p*s > .22). The only exception was a trend for sex for initial correct predictions, $F(1, 9) = 3.424$, $p = .097$, $\eta_p^2 = .276$, reflecting better performance in boys ($M = 5.40$, $SD = 3.91$) compared to girls ($M = 2.13$, $SD = 1.55$; Figure 3). This trend reflects a large effect size for sex on initial correct predictions. Although the difference was not significant, final correct predictions followed the same pattern for boys and girls (*M* for boys = 6.40, $SD = 4.77$, *M* for girls = 3.25, $SD = 3.92$). This finding aligns with the sex differences found in previous tube task studies and other research involving spatial reasoning and young children (for a review, see Halpern & LaMay, 2000). Because this trend was found only for initial but not final correct predictions, the findings indicate that while boys were more likely to make an initially correct prediction, they did not have a higher likelihood of making a final correct prediction than girls after being asked to explain their answers aloud.

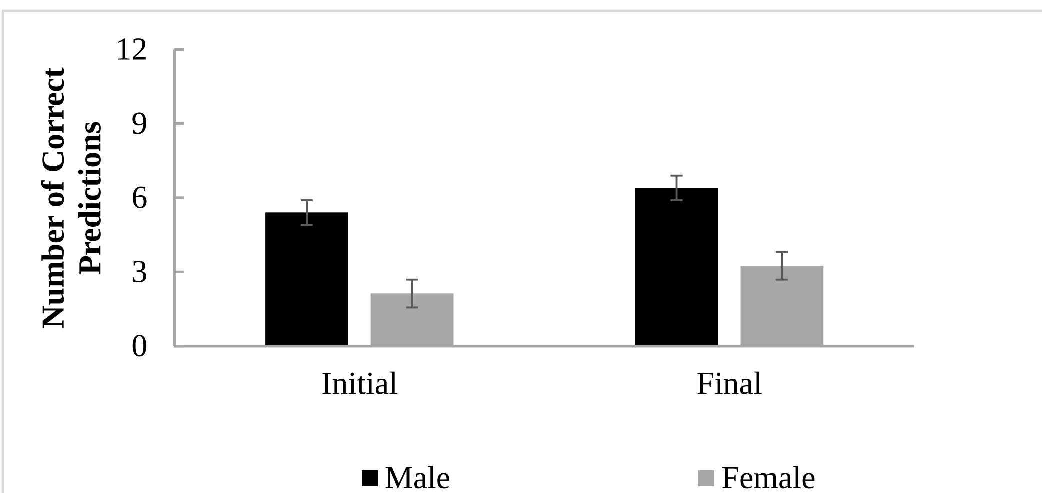

Figure 3. Initial and final correct predictions for males and females in Study 1. Error bars represent mean standard errors.

The explanation for the decrease in sex difference between the initial and final correct predictions may be found in children's switching behaviors. Table 1 describes the number and percentage of trials on which children made an initially correct prediction (row 1) or incorrect prediction (rows 2-6). To understand switching, the trials in which children made an incorrect response (112 out of 156; rows 2-6) are most illustrative. On these trials, children switched to a different answer on 25 trials (22.32% of initially incorrect predictions; row 4), which resulted in a correct prediction on 18 trials (72.00% of switch trials; row 6). In other words, although participants did not switch frequently, when a switch to a different answer did occur, they were more likely to make a correct prediction. Girls made approximately the same number of switches as boys, but boys were also more likely to switch their initially incorrect answers to correct answers. The equivalent amount of switching for girls and boys may explain why there is no trend for sex differences for final correct predictions. Because they had more initially incorrect predictions (row 2), girls switching the same amount as boys narrowed the gap between boys' and girls' performance on the task.

Table 1.

Number of initially correct and incorrect predictions in Study 1 (n males = 5, n females = 8) and Study 2 (n males = 6, n females = 4). Rows 3-4 show percentages of the number of switches or no switches after an initially incorrect prediction. Rows 5-6 show percentages of the number of switches to incorrect or correct predictions after a switch.

		Study 1: Explanation			Study 2: Song		
		Males	Females	Total	Males	Females	Total
1	Initial Correct	45.00%	17.71%	28.21%	25.00%	22.92%	24.17%
2	Initial Incorrect	55.00%	82.29%	71.79%	75.00%	77.08%	75.83%
3	No Switch	75.76%	78.48%	77.68%	74.07%	78.38%	75.82%
4	Switch	24.24%	21.52%	22.32%	25.93%	21.62%	24.18%
5	Switch to Incorrect	12.50%	35.29%	28.00%	21.42%	12.50%	18.18%
6	Switch to Correct	87.50%	64.72%	72.00%	78.57%	87.50%	81.82%

Verbal responses to the question, “Why do you think the ball will come out there?” are recorded below in Table 2. The categories were created based on the common themes described during verbalizations. Participants did not provide a wide variety of words, so all answers are very similar to the example phrases below. Depending on what children said, responses from a single trial may be placed in multiple categories. For example, if a participant said “The ball is going to fall down into the cup,” then the response would be categorized as “Falling,” “Cup/End Point,” and “Ball.” However, answers were rarely this detailed; most answers were generally more vague, such as “It’s going here.” Table 2 shows categories of verbalizations made by participants after being asked why they thought they had made a correct prediction. The categories are rank ordered from most to least frequent. Because single response could be placed in multiple categories, the percentages in the table indicate the percentage of trials in which children made that type of prediction and responded within that category.

Children gave either no response or non-explanatory answers such as “I don’t know” or “just because” on approximately 50% of trials. However, on the remaining trials, they often provided answers related to the path of the ball rolling through the tube, though the most common expressions were still not very specific. For example, most children explained the ball as going from “here to there,” “falling down,” or “going there.” Providing no response or saying “just because” was much more likely to lead to an incorrect prediction. “I don’t know” was only slightly more likely to lead to an incorrect prediction. However, this was one of the most common phrases and may just illustrate that children at this age are unable to fully express their rationale for their predictions. Children who described the ball as going from “here to there” or discussing the cup or end point were equally likely to make a correct or incorrect prediction. Participants who described the ball as “falling” were four times more likely to make an incorrect prediction, potentially because “falling” may refer to the gravity driven response. On the other hand, participants who talked about the ball “going” were much more likely to make a correct prediction. Participants who believed they were in control of the outcome (“participant-controlled”) only made correct predictions, indicating that children who were confident in their answers understood the task well enough to make a correct prediction.

Table 2.

Categories of responses to the "why" question.

Categories	Example Phrase	Percent of Correct Predictions	Percent of Incorrect Predictions	Ratio Correct:Incorrect	Percent of Total Predictions
No response	————	10.34	36.73	1:6	26.92
Here-there	It goes down there then it goes here.	25.86	16.33	1:1	19.87
I don't know	I don't know	15.52	12.24	3:4	13.46
Falling	It's going to fall down.	6.90	16.33	1:4	12.82
Cup/End Point	It's going in the cup.	12.07	7.14	1:1	8.97
"Just because"	Because it will.	6.90	9.18	4:9	8.33
Participant-controlled	I put it there.	20.69	0.00	21:0	7.69
Ball	The ball could drop right here.	12.07	5.10	7:5	7.69
Nonspecific "going"	It's going in that one.	15.52	2.04	9:2	7.05
Questioning	How about there?	1.72	3.06	1:3	2.56
Experimenter-controlled	You're holding the ball here.	1.72	1.02	1:1	1.28
I know	I know it's this one.	1.72	1.02	1:1	1.28
Superstition	You need to do this one 'cause it's the better one.	0.00	1.02	0:1	0.64
Total		37.18	62.82	3:5	100.00

Study 2: Mnemonic Song and Spatial Reasoning

Although providing a verbal explanation did not result in a significantly improved performance on the tube task, children appeared to use certain words consistently. This finding prompted the question of whether it would be beneficial to provide those words to children while asking them to solve the tube task. Providing words that were created by, and thus accessible for, 3-year-olds would provide a different approach to the previous verbal information provided in the studies by Bascandziev and Harris (2010) and Joh et al. (2011). Thus, the mnemonic song for Study 2 was designed to facilitate thought about the connection between the ball and the tubes, and the path the ball takes when dropped down a tube (Figure 2), using some of the key words children used in Study 1. The lyrics are as follows: "I see the ball all the way up there, and I see it roll all the way down here." In participants from Study 1, discussing how the ball was going from "here to there" was among the most common explanations. Furthermore, only when making correct predictions did participants use words such as "see" or "look." The word "roll" was not specifically used by participants, but it refers to the path the ball takes through the tube, rather than it just appearing at the bottom. Importantly, the lyrics did not explicitly tell children the where to find the ball. Instead, the lyrics provided external verbal support in language that is accessible to three-year-olds. They were still required to think about the problem in order to come up with a correct prediction. The lyrics are set to a modified version of "Twinkle Twinkle Little Star" (see Figure 4), which is a familiar melody that is easy for children to sing, learn, and remember. Researchers have found that simple or familiar melodies facilitate better recall (Rainey & Larsen, 2002; Wallace, 1994).

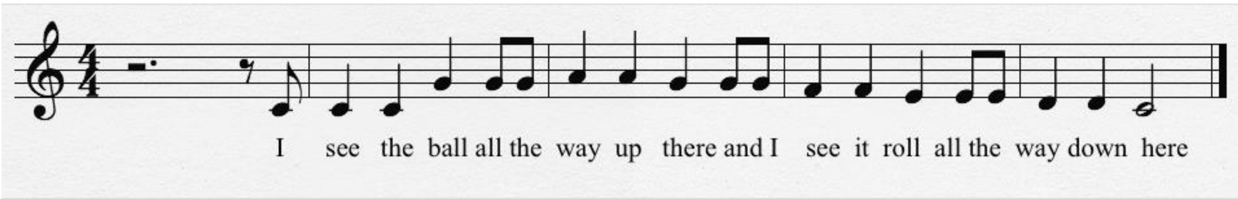

Figure 4. Melody and lyrics of the song used in Study 2.

Pilot testing (4 participants, 3 female, *M* age 45.23 months, *SD* = 5.90) indicated that children at this age, and even children slightly older, were capable of singing along with the experimenter, but had difficulty remembering the lyrics when asked to sing independently. Therefore, the experimenter sang along with the participants to facilitate their usage of the song during the task without exposure to longer amounts of training on the song to reduce the potential for fatigue.

Method

Participants. Nine children between 36 and 42 months of age were recruited using the same methods as Study 1(4 girls and 6 boys; *M* age = 37.07 months, *SD* = 0.75). Children were primarily from middle class families and they were Caucasian (*n* = 7) and biracial (*n* = 3).

Materials and procedure. All materials were procedures were identical to those in Study 1 with the following exceptions. During each phase of the warm up, children heard the experimenter sing the song shown in Figure 4. The procedure for this warm up is shown below in Figure 5. On the test trials, the experimenter held the ball above one of three openings, saying "I'm going to hold the ball here." The experimenter then asked the participants to sing along with her by asking, "Do you want to sing the song with me?" After singing the song together, the experimenter said, "Can you put the ball where you think the ball will come out and tell me when you’re ready?” Once the children indicated that they were ready, the experimenter dropped the ball. This procedure is outlined in the timeline presented in Figure 5.

Familiarization: Tube Introduction	Familiarization: Tube Practice	Familiarization: Cup Introduction	Familiarization: Cup Practice
The experimenter holds up a tube, saying, “Do you see this tube? It’s empty inside. Because it’s empty inside, I can roll a ball through it. And I have a song about the ball rolling through the tube!” The experimenter sings the song then rolls the ball through the tube.	The experimenter asks the participant, “Can you roll the ball back through the tube?” The experimenter catches the ball.	The experimenter holds up the cup, saying, “Do you see this cup? It’s also empty inside. We can use it to catch the ball, and I can sing the song.” The experimenter sings the song then drops the ball into the cup and says, “See? I got it! Do you want to try?”	The experimenter holds the ball above the cup, sings the song, then says, “Tell me when you’re ready for me to drop the ball!” The experimenter drops the ball after the participant says, “Ready!”
Familiarization: Chimney Introduction	Familiarization: Chimney Introduction	Familiarization: Chimney Practice	Familiarization: Chimney Practice
The experimenter shows participants the wooden frame and introduces the chimney openings, pointing to each one and saying, “Do you see these three chimneys up here? And do you see these three chimneys down here? They go with the ball and the cup!”	The experimenter then holds the ball above one of the bottom chimneys and says, “If I hold the ball here, I can sing the song and put the cup where I think the ball is going to come out.” After the experimenter sings the song, she says, “Do you want to try?”	The experimenter holds the ball above one of the bottom chimneys and sings the song, then says, “Can you put the cup where you think the ball is going to come out and tell me when you’re ready?”	After the participant has chosen an opening and said, “Ready!” the experimenter drops the ball. This repeats for all three bottom chimneys.
Familiarization: Single Tube Introduction	Familiarization: Single Tube Introduction	Familiarization: Single Tube Practice	Familiarization: Single Tube Practice
The experimenter attaches one tube to the top left chimney and bottom right chimney, holds the ball above the chimney where the tube is attached.	The experimenter then holds the ball above the top chimney where the tube is attached and says, “If I hold the ball here, I can sing the song and put the cup where I think the ball is going to come out.” After the experimenter sings the song, she says, “Do you want to try?”	The experimenter holds the ball above the chimney where the tube is attached, sings the song, then says, “Can you put the cup where you think the ball is going to come out and tell me when you’re ready?”	After the participant says, “Ready!” the experimenter drops the ball. This repeats for all three tube configurations (top left to bottom right, top middle to bottom left, and top right to bottom middle).

Figure 5. The familiarization phase for Study 2.

Data coding. A primary coder scored each trial from video in the same manner as Study 1, except for verbalizations and gestures. Instead, the coder noted indicators of interest in the song manipulation in the following ways. The coder noted whether the participant *sang along*, to any extent, with the experimenter. Singing along did not include humming; participants must have attempted to sing the lyrics of the song, even if it were only for a portion of the song. Non-verbal responses were coded separately in two categories: whether the participants maintained their *attention* on the experimenter and/or apparatus during the experiment (looking away for two or more seconds was coded as "no") and *moved to the song* (i.e., clapping, dancing). Finally, experimenters coded whether children *responded* yes, no, or did not answer when asked, "Do you want to sing the song with me?"

As in Study 1, the primary coder scored all trials for all participants and a reliability coder scored all trials for 33% of participants. The two coders agreed on 96% of trials for first predictions, 92% of trials for final predictions, 100% of trials for number of switches, 100% of trials for singing along, 100% of trials for attention, 95.8% of trials for moving to the song, and 95.8% of trials for agreeing to sing along. Disagreements were resolved through discussion.

Study Design. As in Study 1, the primary variables of interest were the number of correct initial prediction, correct final predictions, and switching. Trial block and sex were entered as within- and between-subjects variables, respectively. Test age and EVT scores were entered as covariates. Test ages ranged from 36.39 to 38.43 months and EVT scores ranged from 81 to 139 (10^{th} to 99^{th} percentile). Because the score of 81 was considered an outlier, the ANCOVAs were analyzed with and without the outlier (making the range 101 to 139, or 53^{rd} to 99^{th} percentile).

Because removing the outlier made no difference in the results, the findings reported below are from the analyses conducted with the score.

Results and Discussion

On their initial predictions, participants made correct predictions on 24.2% of trials, gravity-driven predictions on 60.8% of trials, and miscellaneous errors on 15% of trials. On their final predictions, participants made correct responses on 37.5% of trials, gravity bias errors on 52.5% of trials, and miscellaneous errors on 10% of trials. These numbers, and pattern of responses, are similar to those from Study 1 and previous work using the tube task (Joh et al., 2011; Joh & Spivey, 2012).

Despite their difficulty, participants actively participated in the task, looking at the experiment 79.2% of the time, agreeing to sing along 45% of the time, verbally singing along 18.3% of the time, and dancing, clapping, or otherwise moving to the song 8.3% of the time.

Again, 2 (trial block) x 2 (sex) mixed-design ANCOVAs were conducted separately for initial and final correct predictions, with test age and EVT scores as covariates. No significant results were obtained from either analysis (all *p*s > .365).

Table 1 displays the switching data for Study 2. After an initially incorrect prediction, children switched to a different answer on 22 of 91 trials (24% of initially incorrect predictions; row 4). Of those switches, 82% led to a correct final prediction. Similar to Study 1, a majority of participants did not switch, but when a switch did occur, participants were likely to choose a correct final prediction. While girls and boys were equally likely to have an initially incorrect prediction and equally likely to make a switch, girls were more likely than boys to switch to a correct answer, although this difference was not significant. This data provides a contrast to the data of Study 1, which showed that not only did girls initially make more incorrect predictions, but also were less likely to switch to a correct prediction than boys.

General Discussion

Two studies were conducted to examine how children think about spatial problems and whether talking aloud (Study 1) or learning a song about the task (Study 2) could help them solve it. Neither study yielded significant results for initial or final correct predictions: spatial reasoning was not related to test age or expressive vocabulary, it did not improve across trials, and it was not significantly related to sex (albeit with some trends).

Why did both manipulations fail to produce significant improvements in spatial ability? One obvious answer is that both studies had small sample sizes; most studies involving the tube task included at least sixteen participants per study (Bascandziev & Harris, 2011; Joh et al., 2011; Joh & Spivey, 2012). But another possible answer may lie in the differences between boys and girls in the two studies.

In Study 1, boys outperformed girls on the tube task, though the difference in the sexes was more pronounced for initial predictions than final predictions. This sex difference is consistent with previous training strategies (Hood, 1995; Joh, under review; Joh & Spivey, 2012). The review of several tube task studies by Joh (under review) indicates that not only do boys make more correct predictions than girls, but they also benefit more from additional information. Furthermore, both spatial working memory and spatial reasoning have been found to be stronger in boys (for a review, see Halpern and LaMay, 2000). However, the smaller difference between the sexes for final predictions in the current study indicates that girls may be at least partially catching up to boys once the explanation strategy was implemented, since initial predictions were made before participants were asked why they thought they had the correct answer. Thus, the explanation manipulation narrowed the gap between the sexes, although boys still did outperform girls as a whole. It has been widely accepted that girls develop language skills earlier than boys (for a review, see Halpern and LaMay, 2000), so even with boys' clear advantage in a spatial task, talking aloud may have helped to diminish this sex difference.

Being asked to explain their answers, a type of social speech, helped narrow the gap between boys and girls from initial to final predictions. On the other hand, in Study 2, expressing the mechanisms of the tube task through song, or even just actively listening to the experimenter sing the song, helped girls improve and allowed them to make approximately the same number of final correct predictions as boys.

In Study 2, girls caught up to boys between initial and final predictions. This is the first tube task study in which girls improved with training more than boys, and arrived at approximately the same number of final correct predictions. Study 2 indicates that a musical or mnemonic manipulation might help girls improve in a way that other interventions have not. Previous research, such as the review by Joh (under review), indicates that boys not only consistently outperform girls, but also benefit from training more than girls. This makes the results in Study 2 that much more striking. Girls and boys made approximately an equal number of initial and final correct predictions. This indicates that unlike most other manipulations, singing a song may help girls just as much as boys. Although the results were not significant, with the limited sample size the trends in the data are still notable.

The difference in the two studies in terms of how sex affects performance may be caused by the way in which the words were presented. Although both studies were meant to utilize expressive vocabulary, in Study 2 participants typically did not sing along and thus received the verbal information, but generally did not express it themselves. Previous research indicates that asking children to talk aloud about a problem may help boys self-regulate more than girls, but utilizing words provided from an external source is easier for girls than boys. For instance, Vallotton & Ayoub (2011) suggest that external verbalization might be more important to 3-year-old boys than girls for self-regulation on tasks. However, Halpern and LaMay (2000) suggest that both the acquisition and usage of verbal information is stronger in girls. Thus, when asked to explain aloud their reasoning for a spatial prediction, boys may benefit more from externalizing their thought processes, as they were asked to in Study 1, whereas girls may benefit more from hearing and utilizing already-provided information, such as the song they hear in Study 2.

Limitations

As previously stated, the nonsignificant results may be due to the small sample size. Furthermore, the sample population is not representative of 3-year-olds as a whole, since participants were primarily from Caucasian, middle class families, and children's EVT scores were primarily above the norm for their age.

In Study 2, it is also possible that the song did not help children make the connection between the task at hand and the solution. Previous research has indicated that during early childhood, children can memorize a song without understanding the meaning behind the song (Calvert, 2001). In Study 2, children were not required to even memorize the song. Many did not sing a long and those that did could not recall the song in its entirety, even with the experimenter singing along. There may need to be a better method for teaching the song, since a majority of children could not or would not sing along. Although the experimenter sang the song eight times during the warm up phase, perhaps this is not enough repetitions to teach children a song.

Both Gfeller (1983) and Calvert and Billingsley (1998) found that rehearsal or exposure of a song over several days was necessary for children to accurately recall the song. When children were asked to recall the song after hearing it only on one day, their recall was highly inaccurate. Thus, perhaps a version of Study 2 that exposes the participants to the mnemonic song for several days before asking them to complete the tube task may enhance their ability to use this song to help them solve the task. It is still undetermined whether being able to actively sing the mnemonic song will help children improve on the task, and this must be explored in future studies.

Future Studies

Further research is needed to determine whether a musical strategy can help children improve on the tube task. To follow up on the potential effects found in Study 2, future research might focus on helping children more thoroughly learn the song, as well as comparing children's performance with and without the song. If participants were brought in for multiple sessions, they may be more likely to thoroughly learn and recite the song. Then, researchers could measure whether the expressive language ability to sing the song plays a role in the outcome. Researchers might also try comparing conditions in which children do not sing the song for trial block 1, and then do sing the song for trial block 2, or vice versa. This method was utilized in the Joh and Spivey (2012) color tube study to determine whether children would improve when a manipulation was added, and whether children would learn from the provided manipulation once it was removed.

Implications

One might apply these findings to spatial reasoning in general, since boys outperforming girls is consistent not only with the tube task but with spatial reasoning skills in general. In Study 1, although boys performed better than girls, the gap in performance was narrowed from initial to final predictions, potentially due to participants being asked to explain their predictions. One might also consider music as a strategy for helping children understand spatial reasoning, or even problem solving in general. Again, the results were not significant, but girls performed just as well as boys, which is unusual on a spatial reasoning task.

It's important to understand what helps, in addition to what does not help, for children and spatial reasoning because it has implications for later math skills, and in turn, STEM career choices (Casey, Nuttall, Pezaris, & Benbow, 1995). Certain subcategories of spatial reasoning ability have been linked to mathematical abilities. Sherman (1979) conducted a study in which spatial visualization was found to be the only other factor besides current mathematical achievement that predicted future success in math classes for high school students. However, another generally strong predictor of math achievement is sex.

Casey, Nuttall, Pezaris, and Benbow (1995) found spatial reasoning, especially mental rotation, to be a mediating factor between gender and performance on the SAT-M (Scholastic Aptitude Test- Math). Boys have been shown to outperform girls in several subcategories of spatial reasoning as well as in math fields and are thus more likely to end up in STEM (Science, Technology, Engineering and Math) fields. With these previous studies in mind, it is crucial to understand how spatial reasoning develops and is nurtured in both boys and girls to maximize the potential for success in mathematics, and in turn, STEM fields. We know from previous research that spatial skills can be taught and are malleable, especially at this young age, so it is important to understand the optimal ways to enhance this skill.

Conclusion

Taken together, the two studies provide further insight into how children think about, and perhaps may even improve, their spatial reasoning abilities. While Study 1 reaffirms the prevalent sex difference in the literature on spatial reasoning, particularly with the tube task, Study 2 indicates that there may be some forms of external information that may benefit girls as well, such as a mnemonic song strategy.

References

Baker, S., Gjersoe, N., Sibielska-Woch, K., Leslie, A., & Hood, B. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science, 14*(2), 270-279. doi:10.1111/j.1467-7687.2010.00972.x

Bascandziev, I., & Harris, P. L. (2010). The role of testimony in young children's solution of a gravity-driven invisible displacement task. *Cognitive Development, 25*(3), 233-246. doi:10.1016/j.cogdev.2010.06.002

Bascandziev, I., & Harris, P. L. (2011). Gravity is not the only ruler for falling events: Young children stop making the gravity error after receiving additional perceptual information about the tubes mechanism. *Journal of Experimental Child Psychology, 109*(4), 468-477. doi:10.1016/j.jecp.2011.03.010

Bilhartz, T. D., Bruhn, R. A., & Olson, J. E. (1999). The effect of early music training on child cognitive development. *Journal of Applied Developmental Psychology, 20*(4), 615-636. doi:10.1016/S0193-3973(99)00033-7

Buchsbaum, D., Bridgers, S., Weisberg, D., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1599), 2202-2212. doi:10.1098/rstb 2012.0122

Calvert, S. L. (2001). Impact of televised songs on children's and young adults' memory of educational content. *Media Psychology, 3*(4), 325-342.

Calvert, S. L., & Billingsley, R. L. (1998). Young children's recitation and comprehension of information presented by songs. *Journal of Applied Developmental Psychology, 19*(1), 97-108. doi:10.1016/S0193-3973(99)80030-6

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*(4), 299-319. doi:10.1016/j.jecp 2004.01.002

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31*(4), 697-705. doi:10.1037/0012-1649.31.4.697

DeLoache, J. S., Cassidy, D. J., & Brown, A. L. (1985). Precursors of mnemonic strategies in very young children's memory. *Child Development, 56*(1), 125-137. doi:10.1111/1467-8624.ep7252820

Freeman, N. H., Hood, B. M., & Meehan, C. (2004). Young children who abandon error behaviourally still have to free themselves mentally: A retrospective test for inhibition in intuitive physics. *Developmental Science, 7*(3), 277-282. doi:10.1111/j.1467-7687.2004.00346.x

Geist, K., & Geist, E. A. (2012). Bridging music neuroscience evidence to music therapy best practice in the early childhood classroom: Implications for using rhythm to increase attention and learning. *Music Therapy Perspectives, 30*(2), 141-144.

Gever, B. E., & Weisberg, R. W. (1970). Spontaneous verbalization of children during problem solving: Effects of task difficulty, age, and social class. *Proceedings Of The Annual Convention Of The American Psychological Association, 5*(Pt. 1), 297-298.

Gfeller, K. E. (1983). Musical mnemonics as an aid to retention with normal and learning disabled students. *Journal of Music Therapy, 20*(4), 179.

Goodman, S. H. (1984). Preschool children's use of speech during phases in task completion. *The Journal of Genetic Psychology: Research and Theory on Human Development, 145*(1),127-132.

Gredler, M. E. (2009). Hiding in plain sight: The stages of mastery/self-regulation in Vygotsky's cultural-historical theory. *Educational Psychologist, 44*(1), 1-19. doi:10.1080/00461520802616259

Habe, K., & Jaušovec, N. (2003). Mozart effect - reality or science fiction? *Psihološka Obzorja / Horizons Of Psychology, 12*(4), 23-32.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*(2), 229.

Hitz, R. (1987). Creative problem solving through music activities. *Young Children, 42*(2), 12-17.

Hood, B. M. (1995). Gravity rules for 2-to 4-year olds? *Cognitive Development*, *10*(4), 577-598. doi:10.1016/0885-2014(95)90027-6

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science, 1*(1), 59.

Hood, B., Carey, S., & Prasada, S. (2000). Predicting the outcomes of physical events: Two-year-olds fail to reveal knowledge of solidity and support. *Child Development*, (6). 1540.

Hood, B. M., Santos, L., & Fieselman, S. (2000). Two-year-olds' naive predictions for horizontal trajectories. *Developmental Science, 3*(3), 328.

Hood, B. M., Wilson, A., & Dyson, S. (2006). The effect of divided attention on inhibiting the gravity error. *Developmental Science, 9*(3), 303-308. doi:10.1111/j.1467-7687.2006.00493.x

Ives, W. (1980). Preschool children's ability to coordinate spatial perspectives through language and pictures. *Child Development*, *51*(4), 1303-1306. doi:10.2307/1129579

Joh, A. S. (under review). Training effects and sex difference in preschoolers' spatial reasoning ability. Submitted to *Developmental Psychobiology*.

Joh, A. S., Jaswal, V. K., & Keen, R. (2011). Imagining a way out of the gravity bias: Preschoolers can visualize the solution to a spatial problem. *Child Development*, *82*(3),744-750. doi:10.1111/j.1467-8624.2011.01584.x

Joh, A. S., & Spivey, L. A. (2012). Colorful success: Preschoolers' use of perceptual color cues to solve a spatial reasoning problem. *Journal of Experimental Child Psychology*, *113*(4),523-534. doi:10.1016/j.jecp.2012.06.012

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences*, *100*(18), 10568-10573.

Kim, I., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science*, *2*(3), 339-362. doi:10.1111/1467-7687.00080

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*(1), 1-33. doi:10.1016/j.cognition.2004.03.009

Manning, B. H., White, C. S., & Daugherty, M. (1994). Young children's private speech as a precursor to metacognitive strategy use during task engagement. *Discourse Processes*, *17*(2), 191-211. doi:10.1080/01638539409544866

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*(11), 1063-1066. doi:10.1111/j.1467-9280.2008.02200.x

Moore, D. S., & Johnson, S. P. (2011). Mental rotation of dynamic, three-dimensional stimuli by 3-month-old infants. *Infancy, 16*(4), 435-445. doi:10.1111/j.1532-7078.2010.00058.x

Moore, K., Peterson, D., O'Shea, G., McIntosh, G., & Thaut, M. (2008). The effectiveness of music as a mnemonic device on recognition memory for people with multiple sclerosis. *Journal of Music Therapy*, *45*(3), 307-329.

Newcombe, N. S., & Frick, A. (2010). Early education for spatial intelligence: Why, what, and how. *Mind, Brain, and Education, 4*(3), 102-111. doi:10.1111/j.1751-228X.2010.01089.x

Örnkloo, H., & von Hofsten, C. (2007). Fitting objects into holes: On the development of spatial cognition skills. *Developmental Psychology*, *43*(2), 404-416. doi:10.1037/0012-1649.43.2.404

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science, 19*(11), 1067-1070. doi:10.1111/j.1467-9280.2008.02201.x

Rainey, D. W., & Larsen, J. D. (2002). The effect of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception: An Interdisciplinary Journal*, (2). 173. doi:10.1525/mp.2002.20.2.173.

Rauscher, F., Shaw, G., & Ky, K. (1995). Listening to Mozart enhances spatial-temporal reasoning: towards a neurophysiological basis. *Neuroscience Letters*, *185*(1), 44-47. doi:10.1016/0304-3940(94)11221-4

Rauscher, F., Shaw, G., Levine, L., Wright, E., Dennis, W., & Newcomb, R. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurological Research, 19*(1), 2-8.

Rauscher, F. H., & Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment. *Early Childhood Research Quarterly, 15*(2), 215-228. doi:10.1016/S0885-2006(00)00050-8

Sherman, J. (1979). Predicting mathematics performance in high school girls and boys. *Journal of Educational Psychology*, *71*(2), 242-249. doi:10.1037/0022-0663.71.2.242

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

Spivey, L. A. & Joh, A. S. (in preparation). Does inhibitory control promote spatial reasoning in preschoolers?

Stiles-Davis, J. (1988). Developmental change in young children's spatial grouping activity. *Developmental Psychology, 24*(4), 522-531. doi:10.1037/0012-1649.24.4.522

Thaut, M. H., Peterson, D. A., McIntosh, G. C., & Hoemberg, V. (2014). Music mnemonics aid verbal memory and induce learning—Related brain plasticity in multiple sclerosis. *Frontiers in Human Neuroscience*, *8.*

Vallotton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly*, 26169-181. doi:10.1016/j.ecresq.2010.09.002

von Hofsten, C. (1980). Predictive reaching for moving objects by human infants. *Journal of Experimental Child Psychology*, 30(3), 369-382. doi:10.1016/0022-0965(80)90043-0

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1471-1485. doi:10.1037/0278-7393.20.6.1471

Williams, K. T. (2007). *Expressive Vocabulary Test* (2^{nd} ed.). Bloomington, MN: NCS Pearson, Inc.