Received: 31 December 2015 | Accepted: 29 June 2016
DOI 10.1002/dev.21445

RESEARCH ARTICLE

WILEY Developmental Psychobiology

Training effects and sex difference in preschoolers' spatial reasoning ability

Amy S. Joh

Department of Psychology, Seton Hall University, South Orange, New Jersey

Correspondence
Amy S. Joh, Department of Psychology, Seton Hall University, 400 South Orange Avenue, South Orange, NJ 07079.
Email: amy.joh@shu.edu

Funding Information
This research was supported by Seton Hall University.

Abstract

This study examined sex difference in spatial reasoning, a type of spatial cognition necessary for everyday activities. An aggregated data set was composed of data from 273 3- to 4-year olds who participated in 12 different studies using variants of the same spatial reasoning task. This data set was used to investigate whether and how sex difference is related to learning opportunities through training. The results showed that boys outperform girls in general, but this sex difference was influenced by training. When children received additional training, boys showed improved spatial reasoning ability compared to girls. But when children did not receive additional training, there was no sex difference. The type and amount of training did not influence the sex difference in this data set. These findings add to our understanding of how sex difference in spatial cognition emerges in early development.

KEYWORDS
preschoolers, sex difference, spatial cognition, spatial development, spatial reasoning, training

1 | INTRODUCTION

It is difficult to overstate Carolyn Rovee-Collier's contributions to developmental research. Rovee-Collier and her collaborators' tireless efforts produced seminal studies on early learning and memory. Their findings have provided an important foundation for what we now know about cognitive development. But *how* Rovee-Collier conducted research has been just as influential as *what* she studied. Rovee-Collier's systematic and methodical examination of memory development, always with an emphasis on developmental trajectories, serves as a reminder of what good research looks like. It also encourages the rest of us to do the same. In this paper, to honor Rovee-Collier's teachings about research, a fundamental topic in early development was investigated in a programmatic manner.

Spatial cognition is a powerful tool that allows us to obtain, process, and make use of information about the location and movement of objects and people in the environment. Spatial cognition refers to an aggregation of related, but unique, cognitive and perceptual skills including the ability to represent the appearance of two- and three-dimensional objects as they move and rotate (mental rotation), perceive the orientation of objects with respect to self (spatial perception), and visualize and manipulate multiple steps of spatial information (spatial visualization) (Halpern & LaMay, 2000; Linn & Petersen, 1985). It includes the ability to represent static, dynamic, intrinsic (intraobject), and extrinsic (interobject) events (Newcombe & Shipley, 2015; Newcombe, Uttal, & Sauter, 2013; Uttal et al., 2013). In short, without spatial cognition, it would be difficult—or impossible—to plan and complete everyday actions such as navigating traffic to cross the street safely, using a map to reach an unfamiliar location, and rearranging the furniture in a room without having to physically move each piece first.

Many studies of spatial cognition have reported a sex difference in which males outperform females. Comprehensive meta-analyses have confirmed this general trend (e.g., Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995) but also revealed that the sex difference can be inconsistent and variable [refer to Alyman and Peters (1993), for a contrary view]. For example, Voyer et al. (1995) examined 286 published studies and found that in 283 cases (99%), there is a male advantage or no sex difference. The effect size for the sex difference was approximately small to medium (Cohen's $d = .37$, $p < .01$). Only three studies reported a female advantage. However, this review also demonstrated that the effect size for sex difference fluctuated across spatial skills and across tasks used to assess the same spatial skill. For example, the sex difference was larger and significant for mental rotation (*d* = .56) but smaller and not statistically significant for spatial visualization (*d* = .19). The sex difference was also significant for spatial perception (*d* = .44), but smaller in size than it was for mental rotation. Within the same spatial category (e.g., mental rotation), the sex difference was more substantial for some tasks (Mental Rotations Test, *d* = .67) than for others (Cards Rotation Test, *d* = .31). Within the same task (e.g., Mental Rotations Test), different methods of scoring (out of 20 vs. out of 40) produced noticeable changes in effect size (*d* = .94 vs. *d* = .70, respectively).

The sex difference in spatial cognition becomes more salient with development. Voyer et al.'s (1995) review showed that for mental rotation, the magnitude of the sex difference in adults (*d* = .66) was double that of children under 13 years of age (*d* = .33). Similarly, for spatial perception, the difference was stronger for adults (*d* = .48) than it was for children (*d* = .33). For spatial visualization, there was no sex difference when studies were collapsed across ages. However, when studies were sorted by age groups, the effect size increased and became significant for adults (*d* = .23), but decreased and remained not significant for children (*d* = .02).

Less is known about the sex difference in very young children. Voyer et al.'s (1995) review included only 74 studies (26% of included studies) with participants under 13 years of age. And of those 74 studies, only a few were conducted on participants under 5 years of age. Since the publication of their review, some studies have shown a sex difference in mental rotation in infants as young as 3 months of age (Moore & Johnson, 2008, 2011; Quinn & Liben, 2008, 2014). But others have not (Hespos & Rochat, 1997; Rochat & Hespos, 1996), suggesting that the sex difference is inconsistent early in life. Indeed, a recent summary of mental rotation studies noted that most infant studies do not show a significant sex difference (Frick, Möhring, & Newcomb, 2014). It appears that the sex difference becomes more consistent around the preschool years (Ehrlich, Levine, & Goldin-Meadow, 2006; Levine, Huttenlocher, Taylor, & Langrock, 1999) and more robust through adolescence and adulthood (Geiser, Lehmann, & Eid, 2008; Johnson & Meade, 1987; Sanders, Soares, & Aquila, 1982).

The sex difference in spatial cognition has far-reaching consequences, notably because earlier spatial skills are related to later performance in STEM (science, technology, engineering, and mathematics) fields. For example, spatial skills at age five can predict math abilities several years later (Gunderson, Ramirez, Beilock, & Levine, 2012). During high school and college, spatial ability is correlated positively with achievement in and preference for math and science courses, and correlated negatively with preference for humanities courses (Lubinski & Benbow, 2006; Shea, Lubinski, & Benbow, 2001). Spatial ability mediates sex differences in SAT math scores (Casey, Nuttall, & Pezaris, 1997). Spatial ability also is a significant predictor for majoring in a STEM subject in college and entering a STEM profession (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Humphreys, Lubinski, & Yao, 1993; Lubinski & Benbow, 2006; Shea et al., 2001; Wai, Lubinski, & Benbow, 2009). Unfortunately, women are underrepresented in STEM fields despite an increase in the overall number of women attending college and graduate school (Ceci, Williams, & Barnett, 2009). When women do work in a STEM field, they tend earn less than their male peers (Olitsky, 2014). The gender gap in STEM achievement is a complex issue with multiple causes and effects. It is possible that one of these causes is related to the sex difference in spatial skills, which emerges early in life.

What causes the sex difference in spatial cognition? Similar to their study of other complex psychological skills, researchers acknowledge that there is a biological component to spatial ability while also emphasizing the interaction between biological and experiential components (e.g., Linn & Petersen, 1985; McGee, 1979). We know, for example, that playing with blocks or art materials—activities that encourage spatial thinking and problem solving—are related to spatial ability (Brosnan, 1998; Caldera et al., 1999). Playing with puzzles is also related to spatial ability, but the relationship is dependent on the quality of puzzle play, and typically, boys show higher quality of puzzle play compared to girls (Levine, Ratliff, Huttenlocher, & Cannon, 2012). Additionally, spatial language is related to spatial ability (Pruden, Levine, & Huttenlocher, 2011): more spatial input from parents leads to more spatial language in children, which leads to better spatial skills later in childhood. Playing with blocks, which demand spatial skills, can increase the use of spatial language (Ferrara, Hirsh-Pasek, Newcombe, Golinkoff, & Lam, 2011), and spatial language can influences the encoding and memory for spatial events (Feist & Gentner, 2007). Training also improves spatial skills (de Acedo Lizarraga & Ganuza, 2003; Law, Pellegrino, & Hunt, 1993; Uttal et al., 2013) and can even decrease the sex difference (Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). Children benefit from a variety of training regimens, including those administered directly or indirectly and over brief or prolonged periods (Baenninger & Newcombe, 1989; Uttal et al., 2013).

Building on previous knowledge, this paper sought to expand the current understanding of the development of sex difference in spatial cognition by examining how experience-related learning opportunities influence children's performance in a spatial task. This paper focused on learning opportunities, rather than biological contributions, with the assumption that it would be more feasible to eventually translate the resulting findings into a plan for promoting spatial skills in young children. Thus, of particular interest was how different types of training opportunities influence spatial performance. To create a data set large and comprehensive enough to detect and examine a possible sex difference, data were aggregated from 12 studies using the same general spatial task. This data set focused on participants between 3 and 4 years of age because this is the developmental period during which the sex difference becomes established in most spatial skills.

1.1 | Assessing spatial reasoning in preschool-age children

This paper focused on spatial reasoning, which refers to the ability to follow and predict the trajectory of moving objects and respond accordingly. Spatial reasoning relies on elements of other spatial skills such as mental rotation, spatial perception, and spatial visualization. However, more so than other spatial skills, spatial reasoning requires attention to and predictions about movement trajectories. As such, spatial reasoning may be particularly important for achievement in fundamental STEM areas such as physics. Yet, spatial reasoning has not been studied as systematically as other spatial skills, resulting in a limited knowledge of how this important skill develops in young children.

Rosenthal (1979) cautioned about the "file-drawer problem" that can plague publications. Studies are typically considered to be publishable if they reveal significant findings; those with non-significant findings are sometimes forgotten and relegated to a file drawer. To mitigate this issue, the current data set included all data, from published and unpublished studies, collected from my lab using the same spatial reasoning task. Thus, the present aggregated data set includes data from studies in which the primary experimental manipulation did not yield significant results. This approach resulted in the inclusion of data from 273 children between three and four years of age, with children tested in 12 studies under various training or control conditions (Table 1).

All children participated in the same general task shown in Figure 1, which is an adaption of Hood's (1995) "chimney" task. A ball is dropped down one of three diagonally intertwined tubes (e.g., location A). The spatial components of the task include anticipating the movement of the ball and placing a cup to catch the ball before it drops. The correct prediction can be derived by simply following the path of the tube into which the ball was dropped (location B). However, children younger than 4 years of age expect the ball to fall straight down (location C) even though this outcome is impossible given the placement of the tubes (Bascandziev & Harris, 2010, 2011; Hood, 1995, 1998; Joh, Jaswal, & Keen, 2011; Joh & Spivey, 2012). These incorrect responses are called "gravity bias errors" (Hood, 1995, 1998) because they suggest that children resorted to a default assumption based on their naïve knowledge of physics—that objects fall vertically due to the effects of gravity. In doing so, children are basing their responses on what they know about how objects typically move

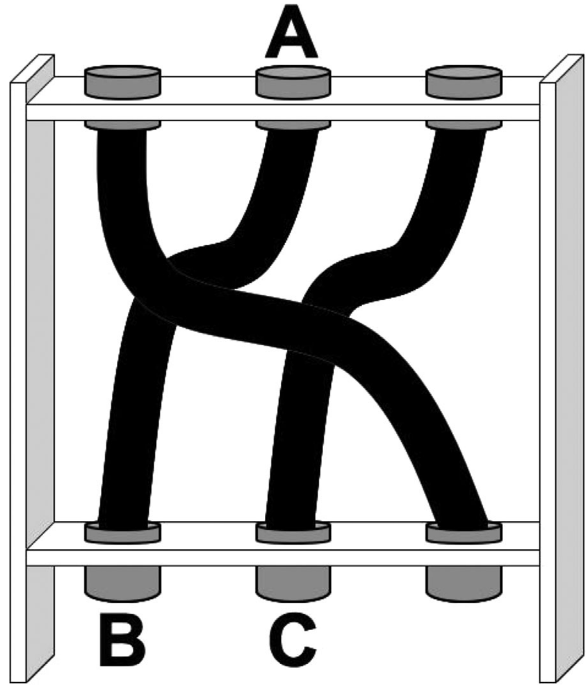

FIGURE 1 Spatial reasoning (SR) task. A ball is dropped down an opening in the top (e.g., A). Location B is the correct response. Location C shows a "gravity-bias error" in which children predict that objects fall straight down regardless of the available path

TABLE 1 Training and control conditions

Training type	Study number	Training level (%)	Test age (months)	Publication	Brief description
Visual imagery	1	100	36–42	Joh et al. (2011)	Visual imagery training (opaque and clear tubes)[a]
	2	100	36–42		Visual imagery training (opaque tubes only)
	3	50	36–42		Visual imagery training (opaque tubes only)
Color cues	4	100	36–42	Joh and Spivey (2012)	Red, yellow, and blue tubes
	5	50	36–42	Joh and Spivey (2012)	Red, yellow, and blue tubes
Motor tracing	6	100	36–42		Tracing training
Verbal explanation	7	100	36–48		Asked to explain prediction
Control (no training)	8	0	36–42	Joh et al. (2011)	Control for words and length of prompt in Study 1
	9	0	36–42	Joh et al. (2011)	General procedure
	10	0	36–42		Visual imagery training during familiarization
	11	0	36–42	Joh and Spivey (2012)	General procedure
	12	0	42–48		General procedure

[a]There was no significant difference in performance between the opaque and clear tube trials.

through space rather than how the ball will move through a tube in this task.

Although the gravity bias is persistent—children make gravity bias errors repeatedly and consistently—it does not show that children do not understand the chimney task. Children succeed in the task if they are presented with a simpler problem with one or two tubes (Hood, 1995) or if the ball moves horizontally or up (not down) through the intertwined tubes (Hood, 1998; Hood, Santos, & Fieselman, 2000). Conversely, older children who no longer make gravity bias errors return to such responses if their attention is taxed with more difficult versions of the task such as avoiding the correct location or keeping tracking of two falling balls (Freeman, Hood, & Meehan, 2004; Hood, Wilson, & Dyson, 2006). Thus, the gravity bias error highlights what is unique and challenging about spatial reasoning: To predict the movement of an object, we must consider our knowledge of object movement as well as ongoing constraints before deciding on a course of relevant action.

Another critical aspect of spatial reasoning, one that makes it particularly informative for furthering our understanding of the sex difference in spatial cognition, is that spatial reasoning can improve with additional learning opportunities. For example, teaching children to "follow that tube with your eyes" or "imagine the ball rolling down the tube" results in fewer gravity bias errors (Bascandziev & Harris, 2010; Joh et al., 2011). Sometimes, simply providing children with more information about the task, such as adding distinctly colored tubes that highlight each path saliently (Joh & Spivey, 2012) or removing the "chimney" pieces from the apparatus (Bascandziev & Harris, 2011), can also lead to improved performance.

In this data set, approximately half of the children received one of the following learning opportunities: (A) prompts to use visual imagery (Studies 1–3 in Table 1); (B) additional perceptual information via colored tubes (Studies 4–5); (C) training to use a motor tracing strategy (Study 6); and (D) prompts to verbally explain the task (Study 7). The remaining children participated in control conditions in which they received no additional training (Studies 8–12). The training conditions are described below.

In addition to training type, training level also varied so that some children received training on 100% of trials (Studies 1, 2, 4, 6, and 7 in Table 1) and others on 50% of trials (the first half of the test trials; Studies 3 and 5). Findings from the 100% groups should show whether participants benefited at all from a particular learning opportunity, whereas findings from the 50% groups should show whether the benefits of a learning opportunity remained after the prompt was removed from the procedure. Participants in the control groups received 0% training.

1.1.1 | Visual imagery

We often use visual imagery by mentally representing events that are not physically present. For example, when rearranging the furniture in a room, we first work through possible configurations mentally to avoid the hassle of having to move heavy furniture more than necessary. To train children to use this helpful strategy, we asked them to "imagine the ball rolling down the tube" at the start of each trial. Children were asked to "imagine" rather than "visualize" or "mentally represent" because of their presumed familiarity with the word "imagine."

Asking children to "imagine the ball rolling down the tube" provides an opportunity for visual imagery, but also an additional reminder about task-relevant components such as "tube" or "roll" as well as additional time to think about it. Therefore, a control group for the visual imagery training study was Study 8 in which children were told that "the ball is going to roll down the bumpy tube." In this way, children received the same time delay through the same instruction containing the same key words, with the exception of the word "imagine."

1.1.2 | Color cues

We also rely on cues from the environment to solve spatial problems. Color is a salient visual cue because it can signal the location of a spatially oriented event (e.g., in an intersection, yellow lines are used to demarcate a crosswalk). To determine whether children are able to extract useful perceptual cues from the environment, and more importantly, whether they can use such cues as the basis for a helpful strategy, children were presented with three tubes of distinct colors. The colors of the tubes (red, yellow, and blue) were chosen because of children's familiarity with primary colors.

1.1.3 | Motor tracing

Sometimes, we also act out solutions to a problem. In the case of the chimney task, physically tracing the path of the relevant tube could provide children with a tangible way to visualize and anticipate the ball's trajectory and help them to understand the significance of the configuration of the tubes. Thus, children in this condition were taught to trace the path of the relevant tube before making a prediction about where they thought the ball would emerge.

1.1.4 | Verbal explanation

Finally, children were prompted to answer the question of "Why do you think the ball will come out of there?" before the start of each trial. Like the visual imagery training, the verbal explanation training was designed to help children to think through the spatial event before making a prediction. However, to encourage children to develop a strategy that eventually leads to independent success, this prompt was more general than the visual imagery prompt. In this way, children could answer—and perhaps learn to use—any strategy that was helpful to them.

1.2 | Current study

There were three aims to the current study. Each aim was exploratory, but each successive aim was designed to build and expand on the outcomes from a previous aim. The first aim was to establish whether there is a sex difference in spatial reasoning. Although some studies using the chimney task reported a significant or marginal sex difference (Hood, 1995; Joh & Spivey, 2012), others did not (Bascandziev & Harris, 2010, 2011; Joh et al., 2011). To examine a possible sex difference, several outcome variables were available for analyses from this aggregated data set including correct predictions and switching behaviors. Correct predictions, not gravity bias errors, were used because they are a more stringent measure of spatial knowledge—not showing a gravity bias error does not mean that participants made a correct prediction. As shown in Table 2, responses were mostly limited to correct predictions and gravity bias errors; children rarely chose the third location. Indeed, across the 12 studies, children averaged just .67 miscellaneous responses out of 12 possible responses (*SD* = .29; range = .25–1.19). Therefore, correct predictions were analyzed to avoid redundant results. Additionally, switching behavior—holding the cup under one opening before moving it to another location—was used to infer children's decision making process. Presumably, more switching would indicate that participants were testing out alternative possibilities.

The second aim was to determine whether there is an effect of training on sex difference in spatial reasoning. In other words, do children benefit from opportunities for learning about spatial reasoning and if so, do boys and girls respond to training in different ways? And finally, the third aim was to examine whether the type and/or amount of training influence the sex difference in spatial reasoning. Do boys and girls benefit differently from specific types of training, or is it simply the opportunity to learn that is most important? The answers to these questions should add to what we currently know about the early differences in spatial skills while highlighting the importance of early opportunities for later outcomes.

2 | METHODS

2.1 | Participants and materials

Two hundred and seventy-three children were tested between 36 and 48 months of age (*M* age = 41.28 months, *SD* = 3.54). The number of boys (*n* = 135) and girls (*n* = 138) was approximately equal, as was the number of participants in the training groups (*n* = 127) and control groups (*n* = 146). Table 2 provides the details for each study.

The chimney apparatus consisted of a wooden frame (62.5 cm high, 59.1 cm wide, and 8.9 cm deep) with three openings at the top and three openings at the bottom (Figure 1). The openings were spaced equally and fitted with a round, white, plastic "chimney" (5.7 cm in diameter and 6.4 cm long). Three opaque, flexible plastic tubes of the same color (approximately 4.4 cm in diameter and 67.6 cm long) could be connected from one of the top chimneys to one at the bottom to create a path for a small ball (2.5 cm in diameter). The ball was made of hard foam and fell down the tubes noiselessly. A small cup was used to catch the ball. [Participants in Studies 1, 8, and 9 were tested with a different frame and tubes, which are described in Joh et al. (2011). The differences were superficial, however, and the overall configurations and use of the apparatus were same across all studies.]

2.2 | General procedure

Participants were tested at a small table with the experimenter sitting or standing across from them. Parents and guardians remained in the testing room with their children but did not provide any answers. Sessions were recorded for later coding.

2.2.1 | Familiarization

Participants were introduced to each component of the task separately, including the tubes, ball, and cup used for catching the ball. First, the experimenter placed a single tube horizontally on the table and said, "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 then rolled a ball through the tube and allowed participants to catch the ball (*horizontal tube familiarization*). This process was repeated, but with the participants rolling the ball through the tube for the experimenter to catch.

Next, the experimenter demonstrated that the cup can be used to catch the ball while simultaneously teaching participants to indicate that they are ready for the ball to be dropped. The experimenter picked up the cup, held a ball above it, and said, "Do you see this cup? It's also empty inside. We can use it to catch the ball. See?" The experimenter then dropped the ball into the cup and exclaimed, "I got it!" (*cup familiarization*). The experimenter then handed the cup to the participants, held a ball above it, and said, "Do you want to try? Tell me when you're ready!" The experimenter dropped the ball only after participants said, "Ready," at this time and also on all subsequent familiarization and test trials.

The experimenter then brought out the wooden frame to demonstrate how the tubes, ball, and cup function in conjunction with the frame. Without any tubes attached to the frame, the experimenter pointed to each of the openings slowly and said, "Do you see these chimneys up here? And do you see these chimneys down here? They go with the ball and the cup." She held the ball over one of the lower openings, a cup under the same opening, and said, "If I hold the ball here, then I can put the cup where I think the ball will come out." After dropping the ball through the chimney and catching it with the cup, she exclaimed, "I got it!" (*chimney familiarization*). Participants were given an opportunity to practice catching the ball dropped through an opening, once at each of the three locations.

Finally, the experimenter brought back the single tube and placed it into the apparatus by connecting it to a top and bottom opening (e.g., top right opening and bottom left opening) to create a diagonal pathway. She held the ball above the top opening, placed the cup below the connected bottom opening, and said, "I'm going to hold the ball here and put the cup where I think the ball will come out." She dropped the ball, caught it with the cup, and exclaimed, "I got it!" (*single tube familiarization*). Participants practiced once at each of the three possible tube positions: top right to bottom left, top middle to bottom right, and top left to bottom middle.

2.2.2 | Test trials

The experimenter inserted all three tubes into the frame to create the configuration shown in Figure 1. Just like she did during familiarization, the experimenter held the ball above one of the three top openings and using the same instructions, asked participants to place the cup under the bottom opening where they thought the ball would come out. Again, the ball was dropped only after participants indicated their readiness. After each trial, the experimenter rotated the frame 180° and switched the opening through which the ball was dropped so that

TABLE 2 Outcome variables by study and sex

		Boys				Girls				All participants			
Training type	Study number	*n*	Correct (*SD*)	Gravity (*SD*)	Above chance	*n*	Correct (*SD*)	Gravity (*SD*)	Above chance	*n*	Correct (*SD*)	Gravity (*SD*)	Above chance
Visual imagery	1	8	8.25 (4.03)	3.13 (3.27)	6	8	6.26 (3.77)	5.63 (3.66)	3	16	7.25 (3.91)	4.38 (3.59)	9
	2	8	7.00 (4.34)	4.50 (4.04)	4	8	1.63 (2.56)	10.13 (2.80)	0	16	4.31 (4.42)	7.31 (4.44)	4
	3	8	5.00 (5.13)	6.38 (5.18)	3	8	3.50 (4.21)	8.38 (4.17)	2	16	4.25 (4.60)	7.38 (4.66)	5
	All visual imagery	24	6.75 (4.53)	4.67 (4.27)	13	24	3.79 (3.93)	8.04 (3.92)	5	48	5.27 (4.46)	6.35 (4.40)	18
Color cues	4	8	9.63 (3.58)	1.25 (1.58)	6	8	8.50 (3.42)	2.50 (2.20)	6	16	9.06 (3.43)	1.88 (1.96)	12
	5	8	7.50 (4.00)	3.88 (3.76)	4	8	4.88 (3.31)	6.25 (2.96)	1	16	6.19 (3.80)	5.06 (3.49)	5
	All color cues	16	8.56 (3.83)	2.56 (3.10)	10	16	6.69 (3.75)	4.38 (3.18)	7	32	7.63 (3.85)	3.47 (3.22)	17
Motor tracing	6	10	8.70 (4.24)	2.90 (3.75)	6	10	3.70 (4.03)	8.20 (4.08)	2	20	6.20 (4.77)	5.55 (4.68)	8
Verbal explanation	7	14	7.21 (4.19)	4.14 (4.00)	7	13	3.54 (4.16)	7.54 (4.14)	2	27	5.44 (4.50)	5.78 (4.34)	9
Control (no training)	8	8	2.88 (3.04)	8.13 (3.27)	1	8	3.13 (1.81)	7.50 (2.56)	0	16	3.00 (2.42)	7.81 (2.86)	1
	9	6	3.17 (4.07)	8.18 (4.07)	1	10	5.1 (4.23)	6.20 (3.85)	3	16	4.38 (4.15)	6.94 (3.92)	4
	10	7	6.29 (3.90)	4.86 (4.18)	2	9	3.67 (4.27)	7.78 (3.83)	2	16	4.81 (4.20)	6.50 (4.13)	4
	11	8	5.13 (4.16)	5.63 (4.03)	3	8	3.13 (3.23)	8.38 (3.78)	1	16	4.13 (3.74)	7.00 (4.03)	4
	12	42	6.50 (4.03)	4.81 (4.05)	20	40	5.65 (3.96)	5.70 (3.96)	15	82	6.09 (4.00)	5.24 (4.00)	35
	All control	71	5.63 (4.07)	5.56 (4.09)	27	75	4.80 (3.86)	6.49 (3.83)	21	146	5.21 (3.97)	6.04 (3.97)	48

Correct, mean number of final correct predictions with standard deviations in parentheses; gravity, mean number of final gravity bias errors with standard deviations in parentheses; above chance, number of participants making more correct predictions than expected by chance (binomial $p < .05$).

participants received a novel ball drop-correct landing location pairing on each trial. This procedure was repeated on 12 test trials.

2.3 | Specific procedures

2.3.1 | Visual imagery (Studies 1 and 8)

In Study 1, the experimenter invited children to use visual imagery during the horizontal tube familiarization by saying, "Have you ever used your imagination? Before I drop the ball, let's imagine the ball as it rolls down the tube." The experimenter stared at the tube for a few seconds to show that she was thinking about the ball in the tube. Then she followed the general procedure with the addition of a prompt to use visual imagery ("Can you imagine the ball rolling down the tube?") during the single tube familiarization and test trials.

In Study 8, participants heard the same key words that could promote spatial reasoning without being told to use visual imagery as in Study 1. During the horizontal tube familiarization, the experimenter asked, "Have you ever played with a tube like this? Before I drop the ball, let's feel all the bumps on this tube." And at the start of the single tube familiarization and test trials, the experimenter stated, "The ball is going to roll down the bumpy tube." This study was analyzed as a control study because as described in the introduction, it was a control for the type and length of the instruction used in Study 1.

2.3.2 | Color cues (Studies 4 and 5)

In Study 4, participants were tested with tubes of three distinct colors—red, blue, and yellow—on all 12 test trials. In Study 5, participants were tested with the colored tubes on the first six test trials only. Before trial 7, the experimenter switched the tubes to three tubes of the same color without drawing any attention to the change. In both studies, the blue tube was used during the horizontal tube and single tube familiarization phases.

2.3.3 | Motor tracing (Study 6)

Participants were trained to follow the path of the relevant tube with their hands. During the horizontal tube familiarization, the experimenter asked, "Can you follow the tube with your hand? Because this is how the ball is going to roll down the tube," and practiced using this motor tracing strategy with participants. The experimenter continued to follow the general procedure but as in the visual imagery study, she encouraged children to use this specific strategy ("Can you follow the tube with your hand?") at the start of the single tube familiarization and test trials.

2.3.4 | Verbal explanation (Study 7)

Participants were asked to explain their prediction before the ball was dropped on the test trials. After participants made a prediction by placing the cup under one of the openings and indicated their readiness, the experimenter asked, "Why do you think the ball will come out there?" The experimenter dropped the ball only after participants provided an answer, even if it was "I don't know." The experimenter did not provide any feedback or corrections to participants' explanations.

2.4 | Outcome variables

All of the studies in this data set were coded for the following on each trial: the top opening into which the ball was dropped, the bottom opening from which the ball emerged, the bottom opening under which participants placed the cup first, the number of times participants switched (held the cup under an opening for at least 2 s without indicating readiness before moving to a different location) and the bottom opening under which participants placed the cup and indicated readiness.

From these codes, it was possible to determine the number of initial correct predictions (total number of trials in which participants made a correct prediction the first time they placed the cup under an opening, with or without indicating readiness), number of final correct predictions (total number of trials in which participants made a correct prediction the final time they placed the cup under an opening and indicated readiness), and number of switches (total number of times participants switched before making a final prediction). For each participant, the minimum possible number of correct predictions was 0 and the maximum was 12. The minimum possible number of switches was also 0 but there was no upper limit to switching. If there were no switches, then the initial correct prediction was the same as the final correct prediction.

From the number of correct predictions, it was possible to determine whether each participant performed better than expected by chance or not. Using a .33 probability of making a correct prediction per trial (because there were three possible prediction locations), participants were categorized as having performed above chance if they made eight or more correct predictions out of the 12 trials (binomial $p < .05$). This outcome variable was applied to both initial and final correct predictions.

3 | RESULTS

ANOVAs were tested for homogeneity of variance through Levene's tests ($ps = .09$). Regression models were tested for error independence with Durbin–Watson statistics (falling between 1.94 and 2.00) and homogeneity of variance with histograms and *p*-plots of residuals (visually scanned for normality). Cohen (1988) provides guidelines for interpreting different types of effect sizes: small ($d = .2$, $w = .1$, $\eta_p^2 = .01$, $R^2 = .01$, and $f^2 = .02$), medium ($d = .5$, $w = .3$, $\eta_p^2 = .06$, $R^2 = .09$, and $f^2 = .15$), and large ($d = .8$, $w = .5$, $\eta_p^2 = .14$, $R^2 = .25$, and $f^2 = .35$). Whenever possible, post hoc statistical power was reported in addition to effect size.

3.1 | Sex difference in spatial reasoning outcome variables (Aim 1)

To determine which outcomes variables, if any, showed a sex difference, separate independent-samples *t* tests were performed on the number of initial correct predictions, final correct predictions, and switches (Bonferroni-adjusted $p = .05/3 = .0167$). Boys outperformed girls on the number of initial correct predictions (boys:

$M = 5.01$, $SD = 3.81$; girls: $M = 3.56$, $SD = 3.34$; $t(271) = 3.35$, $p < .01$, $d = .40$, power = .90) and final correct predictions (boys: $M = 6.57$, $SD = 4.25$; girls: $M = 4.64$, $SD = 3.95$; $t(271) = 3.88$, $p < .01$, $d = .47$, power = .97). In contrast, there was no sex difference in the number of switches (boys: $M = 2.51$, $SD = 2.61$; girls: $M = 2.16$, $SD = 2.37$; $t(271) = 1.16$, $p = .25$, $d = .14$, power = .21). Study-specific results are shown in Table 2.

Chi-squares were conducted on the number of participants performing better than expected by chance. For initial correct predictions, 41/135 boys (30.4%) performed above chance whereas only 22/138 girls (15.9%) did so ($\chi^2(1) = 8.00$, $p < .01$, $w = .17$, power = .81). Likewise, for final correct predictions, more boys (63/135 or 46.6%) performed above chance compared to girls (37/138 or 26.8%; $\chi^2(1) = 11.59$, $p < .01$, $w = .21$, power = .93).

Based on these results, the number of correct predictions and the number of participants performing better than expected by chance were analyzed in the ensuing sections to examine the effects of training on the sex difference in spatial reasoning. Switching was dropped from further analysis. Additionally, because the patterns of findings were identical for initial and final predictions, to avoid redundancy, the upcoming sections focused on final predictions which encompass both initial predictions and switching. (Note: All analyses conducted on final correction predictions were also conducted on initial correct predictions, which showed the same patterns of findings.)

3.2 | Effects of training on sex difference in spatial reasoning (Aim 2)

To determine whether training influences spatial reasoning, either alone or in conjunction with sex, a 2 (training: any or none) × 2 (sex: male or female) ANOVA was conducted on the number of final correct predictions. Because there was a positive correlation between test age and the number of final correct predictions ($r(273) = .22$, $p < .01$), test age was entered as a covariate to control for the effects of age. As expected, age was a significant covariate ($F(1, 268) = 19.64$, $p < .01$, $\eta_p^2 = .09$, power = .99). The ANOVA also showed main effects of training ($F(1, 268) = 10.78$, $p < .01$, $\eta_p^2 = .04$, power = .91) and sex ($F(1, 268) = 13.86$, $p < .01$, $\eta_p^2 = .05$, power = .96), as well as an interaction between the two variables ($F(1, 268) = 5.47$, $p = .02$, $\eta_p^2 = .02$, power = .64). As shown in Figure 2, the interaction was due to the boys outperforming the girls with training (boys: $M = 7.61$, $SD = 4.23$; girls: $M = 4.46$, $SD = 4.07$; $t(125) = 4.27$, $p < .01$, $d = .76$, power = .99), but not without it (boys: $M = 5.63$, $SD = 4.07$; girls: $M = 4.80$, $SD = 3.86$; $t(144) = 1.20$, $p = .21$, $d = .21$, power = .24). Put another way, while the boys improved with training ($t(133) = 2.76$, $p < .01$, $d = .48$, power = .79), the girls did not ($t(136) = 50$, $p = .62$, $d = .09$, power = .08).

To investigate which variable, if any, exerted a stronger influence on spatial reasoning, a hierarchical multiple regression model was built with age entered into block 1 as a control variable. Training and sex were entered into block 2 as predictors. This model was able to explain a total of 11.6% of the variance in the number of final correct predictions ($F(3, 269) = 12.93$, $p < .01$; block 2 predictors: $f^2 = .13$,

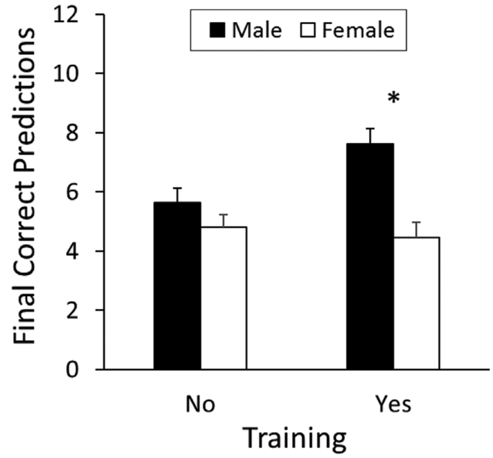

FIGURE 2 Mean number of final correct predictions for each training type by sex. No training = Studies 8–12. Yes training = Studies 1–7. Error bars represent mean standard errors. *$p < .01$

power = .99). As shown in Table 3, training and sex were equally strong predictors of spatial reasoning performance.

Finally, a hierarchical logistic regression model was used to examine the predictors for whether children performed better than expected by chance. Like the multiple regression model, age was entered into block 1 as a control and training and sex were entered into block 2 as predictors. This model differed from the multiple regression model, however, because the criterion was defined as the likelihood that a participant would perform better than expected by chance. This model was significant ($\chi^2(3) = 28.96$, $p < .01$; Cox and Snell $R^2 = .10$, Nagelkerke $R^2 = .14$) and as shown in Table 4, both training and sex were significant predictors.

3.3 | Effects of type or level of training on sex difference in spatial reasoning (Aim 3)

To determine whether the type or level of training influences the sex difference in spatial reasoning, analyses in this section were conducted only on participants who received training. A 4 (training type: visual imagery, color cues, motor tracing, or verbal explanation) × 2 (training level: 100% or 50%) × 2 (sex: male or female) ANOVA was conducted on the number of final correction predictions. In contrast to the analyses conducted for Aim 2, when the data from the control groups were removed for Aim 3, age was not correlated with the number of

TABLE 3 Summary of hierarchical multiple regression model predicting the number of final correct predictions for all participants ($n = 273$)

Variable	R^2 change	B	SE	β	p
Block 1	.05*				
Age		.26	.07	.22	.000
Block 2	.08*				
Training		−1.70	.48	−.20	.000
Sex		1.69	.52	.20	.001

*$p < .01$.

TABLE 4 Summary of hierarchical logistic regression model predicting the probability that a participant performed better than expected by chance for all participants (n = 273)

Variable	*B*	*SE*	Wald	Odds ratio	*p*
Block 1					
Age	.165	.04	14.06	1.18	.000
Block 2					
Training	.85	.30	7.72	2.33	.005
Sex	−.82	.267	9.42	.44	.002

correct final predictions (r(127) = .12, p = .19). Therefore, age was not entered as a covariate in this ANOVA. This ANOVA revealed main effects of training type (F(3, 114) = 3.89, p = .01, η_p^2 = .09, power = .81), training level (F(1, 114) = 4.87, p = .03, η_p^2 = .04, power = .59), and sex (F(1, 114) = 12.53, p < .01, η_p^2 = .10, power = .94). However, there were no interactions (ps > .32). Pairwise comparisons showed that the main effect of training type was due to the children in the color cue studies (M = 7.63, SD = 3.85, n = 32) performing better than children in the visual imagery studies (M = 5.27, SD = 4.46, n = 48; p = .04, d = .57, power = .69). None of the other comparisons were significant (ps > .11). The main effect of training level was due to a higher performance in the 100% groups (M = 6.33, SD = 4.47, n = 95) compared to the 50% groups (M = 5.22, SD = 4.26, n = 32, d = .25, power = .23). And, the main effect of sex was due to boys (M = 7.61, SD = 4.23, n = 64) outperforming the girls (M = 4.46, SD = 4.07, n = 63, d = .72, power = .98).

As before, a linear regression model was used to determine the contribution of training type and level on the number of final correct predictions. Unlike in Aim 2, age was removed from this model because it was not correlated with performance. Training type (dummy coded into m-1 groups), training level, and sex were entered as predictors. This model accounted for 17.7% of the variance in the number of final correct predictions (F(5, 121) = 6.43, p < .01, f^2 = .22, power = .99). Table 5 shows that most of the influence stems from being a boy, color cues training, and training on 100% of trials, in that order.

Again, a logistic regression model was created to examine the probability that children performed better than expected by chance. The predictor variables were identical to the multiple regression model above. This model was also significant ($\chi^2(5)$ = 21.21, p < .01; Cox and Snell R^2 = .15, Nagelkerke R^2 = .21) and showed similar outcomes as the multiple regression model (Table 6).

TABLE 5 Summary of multiple regression model predicting the number of final correct predictions for participants who received training (n = 127)

Variable	*B*	*SE*	*β*	*p*
Training type: visual imagery	.59	1.02	.07	.564
Training type: color cues	3.29	1.15	.32	.005
Training type: motor tracing	.81	1.19	.07	.494
Training level	2.11	.93	.21	.025
Sex	−3.16	.71	−.36	.000

TABLE 6 Summary of logistic regression model predicting the probability that a participant performed better than expected by chance for participants who received training (n = 127)

Variable	*B*	*SE*	Wald	Odds ratio	*p*
Training type: visual imagery	.59	.56	1.12	1.80	.291
Training type: color cues	1.55	.65	5.70	4.72	.017
Training type: motor tracing	.35	.65	.29	1.42	.589
Training level	1.18	.54	4.85	3.26	.028
Sex	−1.44	.40	12.65	.24	.000

4 | DISCUSSION

4.1 | Detecting a male advantage in preschoolers' spatial reasoning

An aggregated data set was created to examine exploratory questions regarding whether and how training-related learning experiences influence the sex difference in spatial reasoning. In the specific context of spatial performance, these questions were designed to expand our current understanding of spatial development. In the broader context of cognitive development, the questions were posed with the anticipation that in the future, we may use the findings to promote children's spatial skills within and across domains. This idea, akin to Barnett and Ceci's (2002) notion of far transfer, has practical implications for crucial developmental outcomes such as education and career achievements.

The results showed that there is a sex difference in spatial reasoning in preschool-age children. Children mostly made correct predictions or gravity bias errors, rarely selecting the third miscellaneous location. Therefore, responses were assessed in several different ways, including the mean number of trials in which participants made correct predictions as well as the total number of children who performed better than expected by chance. Overall, boys performed better than girls on both types of measures, confirming that the male advantage was not due to a few high performers exaggerating the sex difference. Rather, the sex difference was generalized such that each boy, on average, succeeded more frequently than each girl, and the total number of boys who performed well was greater than that of girls.

There was no sex difference in switching behaviors because boys and girls switched equally frequently. This finding, along with the interaction between sex and training, demonstrates that boys outperformed girls right off the bat. It was not the case that boys and girls started out similarly but boys ended up with more correct predictions due to learning (e.g., trying out other solutions before selecting a final correct prediction or improving over trials in the training studies). Rather, from the first trials, boys quickly made use of additional information provided in the training studies, producing a higher number of correct predictions. In contrast, girls continued to commit gravity bias errors in trial after trial, as if the training conditions did not offer additional or helpful information for spatial reasoning.

For the number of final correct predictions, the effect size for the sex difference was medium (d = .47). This effect size is larger than what was found by Voyer et al., 1995 for children under 13 years of age for mental rotation, spatial perception, and spatial visualization. In fact, this effect size is similar in magnitude to adults' spatial perception, which was the skill with the second strongest effect size for adults. Yet, some of the past studies using the same chimney task—including one from my own lab—failed to detect a sex difference (Bascandziev & Harris, 2010, 2011; Joh et al., 2011). Why is there such an inconsistency, especially in the developmental literature? One possibility is the file-drawer problem described in the introduction. In most spatial cognition research, sex difference is not the main focus of a study; it is typically a byproduct—sometimes, an unwelcome one—resulting from the main experimental manipulation. Thus, as warned by Rosenthal (1979), it may be that some studies with a sex difference are not published.

Another possible explanation for the inconsistency is that previous studies were conducted with relatively small samples. For example, Hood's (1998) study was conducted on 15 children. Bascandziev and Harris's (2010) study was conducted with nine children per condition in each age group; their 2011 study was conducted with 20 children per condition. In the studies conducted in my lab, the most typical sample size was 16 children (Table 2); the largest sample size of 82 in Study 12 was an exception, not the rule. This is a methodological issue that is familiar to researchers working with special populations such as infants and children. Some of the most influential developmental studies were conducted with small samples, such as McGraw's detailed investigation of motor development through the co-twin method (McGraw, 1935), Piaget's theory of cognitive development based on observations of his three children (Piaget, 1952), and more recently, Rovee-Collier and co-workers programmatic body of work on the development of infant memory (e.g., Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; n = 6).

An advantage of using small sample sizes, of course, is that researchers are able to gain access to and study a population that is challenging to recruit. But an important disadvantage of relying on small sample sizes is that we may miss some findings, especially those that are unstable because they are newly developing or changing with development. Consequently, it may be that the sex difference in spatial reasoning, or more broadly, in spatial cognition, is difficult to detect with small sample sizes because it is influenced by the constraints of early development. It may be that only large data sets such as the one aggregated for this paper are capable of detecting this phenomenon more reliably. For example, Hood (1995) also found a sex difference in his Experiment 1—which included 209 children. With 273 participants in this paper, when findings were significant, post hoc power averaged .84 (SD = .20). However, individual values varied across analyses (range = .23–.99; median = .92). Tellingly, power was higher when all or most participants were included in an analysis as they were for Aim 1 (M = .90, SD = .07, median = .92) and Aim 2 (M = .90, SD = .13, median = .96). Power tended to be lower for the analyses in Aim 3 (M = .75, SD = .27, median = .81), which included a smaller subset of participants.

4.2 | Explaining training effects on sex difference in spatial reasoning

The results from the current study also showed that the sex difference is dependent on whether or not children receive additional learning opportunities through training. Not only was there an interaction between the two variables, each variable was a roughly equally useful predictor of the outcome variables. However, neither the type nor level of training influenced the sex difference. It was expected that training would lead to an improvement in spatial reasoning. After all, a number of previous studies have documented the benefits of training on other spatial skills (Uttal et al., 2013). But the current findings raise the question of how training influences spatial reasoning and why boys benefit more from training than girls.

Perhaps training promoted spatial reasoning because the additional learning opportunities changed the manner in which relevant information was weighted, combined, and used during the task. According to Newcombe and Huttenlocher's (2006) adaptive combination theory, spatial knowledge is constructed from a *combination* of different sources of information. The resulting combination is *adaptive* because information is selected in a weighted, Bayesian fashion. In this account, experience is particularly crucial for development because it can change the weight of an information source. For example, Ratliff and Newcombe (2008) tested adults in rooms of different sizes (small or large) with access to both geometric cues (size and shape of walls in the room) and featural cues (salient landmark of colorful fabric on a white wall). They found that cue preference changed with room size such that participants favored featural cues in the large room and geometric cues in the small room. However, cue preference also changed as a function of trial number and size of the initial training room, showing that participants were actively encoding and using a combination of different cues, with the weights of cues changing with experience.

Children's responses in our spatial reasoning task suggest that similar principles are at work. Children may be weighting different pieces of information such as what they know about how an object typically moves through space, how the tubes in the chimney task constrain such movement, and other task-related information such as whether and how the addition of multiple tubes change the anticipated trajectory of the ball. They then combine and select what they deem to be most reliable, valid, and salient information for the current problem. For children in the control groups, this process may mean going with what they know about typical object movement. They assume that like most things, the ball in the chimney task will fall straight down—leading to a gravity bias error. In contrast, for children in the training groups, this process may mean an improved ability to attend to information that helps them to predict the movement of the ball. For example, the presentation of colored tubes may highlight the diagonal pathway as well as the impossibility of a vertical drop regardless of the number of tubes—leading to a correct prediction. Thus, through training, children learn to use the most adaptive combination of information for the spatial task.

But why do boys benefit more from training experience than girls? Table 2 shows that when participants received training (Studies 1–7), boys outperformed girls in every single study. However, when participants did not receive training (Studies 8–12), the pattern changed: boys outperformed girls in three studies, and girls outperformed boys in two studies. Taken together, these findings suggest that the sex difference in spatial reasoning may arise from a difference in the ability to take advantage of learning opportunities. And these learning opportunities are important and pervasive. They include general and indirect opportunities, such as a home environment that is more "boy-friendly" or "girl-friendly" (Pomerleau, Bolduc, Malcuit, & Cossette, 1990). They also include more specific and direct opportunities, such as access to toys and exposure to play opportunities that promote spatial thinking. As noted in the introduction, playing with puzzles or blocks is related to spatial development. These are also activities that are traditionally marketed toward boys. For example, blocks designed for young children come in themed sets such as fire stations, super heroes, train stations, and camping trips. Clearly, the default audience is boys. Block sets for girls are treated as exceptions, with "girl-friendly" themes such as taking care of babies and setting up cafes. Given these early experiences, it is not surprising that by the time children enter school, they show marked differences in toy preference (Liss, 1981). And these differences—in experiences and preferences—combined with biology, may predispose boys to seek and benefit more from spatially oriented learning experiences.

4.3 | Future directions

Carolyn Rovee-Collier dedicated her research career to working out the principles of early learning and memory. This endeavor involved numerous collaborators, participants, questions, grants, papers, and years. Likewise, although our knowledge of spatial cognition continues to grow, much work remains before we can take our current knowledge and apply it in ways that can promote spatial skills in children. First, we need to figure out the most effective way—or ways, if boys and girls respond differently to training methods—of encouraging and supporting spatial thinking in children. In the current study, color cues were shown to be most effective for producing correct responses. Unfortunately, this is the type of training that is least likely to lead to a generalized structural change in spatial thinking. For one, the benefits of using cues from the environment dissipate as soon as the cues disappear. For another, this training does not encourage children to figure out the answer on their own; it merely gives them the answer. The other training methods—using visual imagery, tracing the path, and explaining the solution—are more likely to encourage children to figure out the answer on their own. However, though they did produce some improvements over the control conditions, these training methods were less effective for the preschoolers in this data set. Thus, further study is required to identify more useful training methods for young children.

Second, we must investigate longer term effects of training on the sex difference in spatial reasoning. All of the spatial reasoning studies discussed in this paper, including the chimney studies conducted in other labs, examined the immediate effects of training. As such, the only conclusion that can be drawn with regard to training and sex difference is that of short-term effects. But what about long-term effects? Can brief training lead to durable changes in spatial reasoning? Does the sex difference from training hold over time? What does the developmental trajectory look like for boys and girls who do or do not receive training? Can girls catch up to boys in the long-run if they receive extensive training?

Finally, we must place spatial reasoning in a broader context—both for spatial cognition and for STEM-related skills. Spatial reasoning is unique in that it appears to encompass the other aspects of spatial abilities such as mental rotation and spatial visualization. Where and how spatial reasoning should be categorized within the broader domain of spatial cognition, and how spatial reasoning skills are related to the other spatial abilities, is not yet clear. More broadly, it seems likely that spatial reasoning should be related to performance in STEM fields given its emphasis on logical reasoning about how objects move in space. Still, its precise relationship to STEM fields has yet to be established. The continuation of current work in these directions should help researchers support the development of important spatial skills in children.

ACKNOWLEDGMENTS

I am grateful for the guidance and encouragement of my mentors, including my first mentor, Carolyn Rovee-Collier. I would like to thank my students and lab managers for their enthusiastic assistance with participant recruitment, data collections, and data coding. Finally, I would like to thank the hundreds of children and their parents who volunteered their time generously.

This work was supported by the University Research Council Summer Stipend from Seton Hall University.

REFERENCES

Alyman, C., & Peters, M. (1993). Performance of male and female children, adolescents and adults on spatial tasks that involve everyday objects and settings. *Canadian Journal of Experimental Psychology*, *47*, 730–747. Available online at: http://doi.org/10.1037/h0078874

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*, 327–344. Available online at: http://doi.org/0360-0025/89/0300-0327506.00/0

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637. Available online at: http://doi.org/10.1037//0033-2909.128.4.612

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*, 233–246. Available online at: http://doi.org/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*, 468–477. Available online at: http://doi.org/10.1016/j.jecp.2011.03.010

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science*, *11*, 474–480. Available online at: http://doi.org/10.1111/1467-9280.00291

Brosnan, M. J. (1998). Spatial ability in children's play with Lego blocks. *Perceptual and Motor Skills*, *87*, 19–28. Available online at: http://doi.org/10.2466/pms.1998.87.1.19

Caldera, Y. M., McDonald Culp, A., O'Brien, M., Truglio, R. T., Alvarez, M., & Huston, A. C. (1999). Children's play preferences, construction play with blocks, and visual-spatial skills: Are they related? *International Journal of Behavioral Development*, *23*, 855–872. Available online at: http://doi.org/10.1080/016502599383577

Casey, B. M., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: a comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, *33*, 669–680. Available online at: http://doi.org/10.1037/0012-1649.33.4.669

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*, 218–261. Available online at: http://doi.org/10.1037/a0014412

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsadale, NJ: Erlbaum.

de Acedo Lizarraga, M. S., & Ganuza, J. G. (2003). Improvement of mental rotation in girls and boys. *Sex Roles*, *49*, 277–286. Available online at: http://doi.org/0360-0025/03/0900-0277/0

Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, *42*, 1259–1268. Available online at: http://doi.org/10.1037/0012-1649.42.6.1259

Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, *35*, 283–296.

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block talk: Spatial language during block play. *Mind, Brain and Education*, *5*, 143–151. Available online at: http://doi.org/10.1111/j.1751–228X.2011.01122.x

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*, 277–282. Available online at: http://doi.org/10.1111/j.1467–7687.2004.00346.x

Frick, A., Möhring, W., & Newcomb, N. S. (2014). Development of mental transformation abilities. *Trends in Cognitive Sciences*, *18*, 536–542. Available online at: http://doi.org/10.1016/j.tics.2014.05.011

Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence*, *36*, 556–563. Available online at: http://doi.org/10.1016/j.intell.2007.12.003

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, *48*, 1229–1241. Available online at: http://doi.org/10.1037/a0027433

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, *12*, 229–246. Available online at: http://doi.org/10.1023/A:1009027516424

Hespos, S. J., & Rochat, P. (1997). Dynamic mental representation in infancy. *Cognition*, *64*, 153–188. Available online at: http://doi.org/10.1037/0033-295X.94.4.427

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

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science*, *1*, 59–63. Available online at: http://doi.org/10.1111/1467-7687.00013

Hood, B. M., Santos, L., & Fieselman, S. (2000). Two-year-olds' naive predictions for horizontal trajectories. *Developmental Science*, *3*, 328–332. Available online at: http://doi.org/10.1111/1467-7687.00127

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

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology*, *78*, 250–261. Available online at: http://doi.org/10.1037/0021-9010.78.2.250

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*, 744–750. Available online at: http://doi.org/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*, 523–534. Available online at: http://doi.org/10.1016/j.jecp.2012.06.012

Johnson, E. S., & Meade, A. C. (1987). Developmental patterns of spatial ability: An early sex difference. *Child Development*, *58*, 725–740. Available online at: http://doi.org/10.2307/1130210

Law, D. J., Pellegrino, J. W., & Hunt, E. B. (1993). Comparing the tortoise and the hare: Gender differences and experience in dynamic spatial reasoning tasks. *Psychological Science*, *4*, 35–40. Available online at: http://doi.org/10.1111/j.1467–9280.1993.tb00553.x

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*, 940–949. Available online at: http://doi.org/10.1037/0012-1649.35.4.940

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology*, *48*, 530–542. Available online at: http://doi.org/10.1037/a0025913

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498. Available online at: http://doi.org/10.2307/1130467

Liss, M. B. (1981). Patterns of toy play: An analysis of sex differences. *Sex Roles*, *7*, 1143–1150.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*, 316–345. Available online at: http://doi.org/10.1111/j.1745–6916.2006.00019.x

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, *86*, 889–918. Available online at: http://doi.org/10.1037/0033-2909.86.5.889

McGraw, M. B. (1935). *Growth: A study of Johnny and Jimmy*. New York, NY: D. Appleton-Century Co.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*, 1063–1066. Available online at: http://doi.org/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*, 435–445. Available online at: http://doi.org/10.1111/j.1532–7078.2010.00058.x

Newcombe, N. S., & Huttenlocher, J. (2006). Development of spatial cognition. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology* (Vol. 2. Cognition, perception and language, 6th ed., pp. 734–776). Hoboken, NJ: John Wiley & Sons.

Newcombe, N. S., & Shipley, T. F. (2015). Thinking about spatial thinking: New typology, new assessments. In J. S. Gero (Ed.), *Studying visual and* *spatial reasoning for design creativity* (pp. 179–192). New York, NY: Springer.

Newcombe, N. S., Uttal, D. H., & Sauter, M. (2013). Spatial development. In P. D. Zelazo (Ed.), *Oxford handbook of developmental psychology* (Vol. 1: Body and mind, pp. 564–590). New York: Oxford University Press.

Olitsky, N. H. (2014). How do academic achievement and gender affect the earnings of STEM majors? A propensity score matching approach. *Research in Higher Education*, *55*, 245–271. Available online at: http://doi.org/10.1007/s11162-013-9310-y

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International University Press.

Pomerleau, A., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first two years of life. *Sex Roles*, *22*, 359–367.

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. Available online at: http://doi.org/10.1111/j.1467–7687.2011.01088.x

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

Quinn, P. C., & Liben, L. S. (2014). A sex difference in mental rotation in infants: Convergent evidence. *Infancy*, *19*, 103–116. Available online at: http://doi.org/10.1111/infa.12033

Ratliff, K. R., & Newcombe, N. S. (2008). Reorienting when cues conflict: Evidence for an adaptive-combination view. *Psychological Science*, *19*, 1301–1307. Available online at: http://doi.org/10.1111/j.1467–9280.2008.02239.x

Rochat, P., & Hespos, S. J. (1996). Tracking and anticipation of invisible spatial transformations by 4-to 8-month-old infants. *Cognitive Development*, *11*, 3–17. Available online at: http://doi.org/10.1016/S0885-2014(96)90025-8

Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. Available online at: http://doi.org/10.1037/0033-2909.86.3.638

Rovee-Collier, C. K., Sullivan, M. W., Enright, M., Lucas, D., & Fagen, J. W. (1980). Reactivation of infant memory. *Science*, *208*, 1159–1161.

Sanders, B., Soares, M. P., & Aquila D’, J. M. (1982). The sex difference on one test of spatial visualization: A nontrivial difference. *Child Development*, *53*, 1106–1110. Available online at: http://doi.org/10.2307/1129153

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*, 604–614. Available online at: http://doi.org/10.I037//0022-0663.93.3.604

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies.. *Psychological Bulletin*, *139*, 352–402. Available online at: http://doi.org/10.1037/a0028446

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

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

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review*, *15*, 763–771. Available online at: http://doi.org/10.3758/PBR.15.4.763