American Psychological Association

Journal of Experimental Psychology: General

ISSN: 0096-3445

2020, Vol. 149, No. 6, 1116–1136
http://dx.doi.org/10.1037/xge0000699

Children's Exposure to Spatial Language Promotes Their Spatial Thinking

Marianella Casasola, Wendy S. Wei, Daniel D. Suh, Patricia Donskoy, and Ashley Ransom
Cornell University

Does spatial language contribute to the growth of preschool children's spatial skills? Four-year-old children ($N = 50$) were randomly assigned to a play-only ($n = 24$) or a spatial-language and play condition ($n = 26$). Their mental rotation and spatial vocabulary were assessed at baseline and several days after 5 play sessions. Children in the spatial-language condition scored higher at posttest on a mental rotation task than those in the play-only condition. The amount and diversity of experimenter spatial language during the play sessions accounted for a significant amount of the variance on children's posttest mental rotation. Significant gains in mental rotation were replicated in a second study ($N = 34$) with a broader range of play activities and with children enrolled in Head Start. These results show that the facilitative effects of spatial language on spatial cognition are not restricted to the context in which the spatial language is provided. In particular, 4-year-old children's experience with spatial language during play can transfer to promote their mental rotation.

Keywords: mental rotation, spatial cognition, spatial language, spatial play

Spatial skills, the ability to encode, remember, and mentally manipulate the spatial features and relations of objects or space, are central to our daily functioning. They allow us to recall the location of car keys, navigate a route to work, and provide directions to the nearest café. Spatial skills also are pivotal for academic achievement (e.g., Newcombe, 2010). They predict later math competence (e.g., Lauer & Lourenco, 2016; Mix et al., 2016; Verdine, Irwin, Golinkoff, & Hirsh-Pasek, 2014) and entry into the Science, Technology, Engineering and Mathematics (STEM) fields (e.g., Wai, Lubinski, & Benbow, 2009). Despite individual differences in spatial performance, evident even in infancy (e.g., Lauer & Lourenco, 2016; Moore & Johnson, 2008, 2011; Quinn & Liben, 2008, 2014), spatial skills improve with training (Baenninger & Newcombe, 1989; Terlecki, Newcombe, & Little, 2008; Uttal et al., 2013; Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). Across two experimental studies, we test one pathway for promoting 4-year-old children's spatial skills: increasing their exposure to spatial language, terms that refer to the spatial features of objects, such as their shape (e.g., "triangle"), dimensions (e.g., "big"), and orientation (e.g., "upside down"), as well as to the location and spatial relations of objects (e.g., "in the corner," "below"). The present studies break new ground by probing the scope and durability of spatial language's facilitative effect on preschool children's spatial skills. In particular, we seek to establish a causal link between increasing children's exposure to spatial language during a short-term training and advances in their spatial thinking.

A number of theoretical views attribute a central role to spatial language in the development of spatial cognition. Spatial language is argued to facilitate abstraction of relational commonalities (Casasola, 2005; Loewenstein & Gentner, 2005; Xu, 2016), reduce cognitive processing (Pruden, Levine, & Huttenlocher, 2011), as well as strengthen, and even restructure, early spatial representations (e.g., Carey, 2009; Gentner, Ozyürek, Gürcanli, & Goldin-Meadow, 2013; Hermer-Vazquez, Moffet, & Munkholm, 2001; Pyers, Shusterman, Senghas, Spelke, & Emmorey, 2010). These claims are well supported by experimental studies. Providing spatial labels enhances children's visual search, reorientation, spatial reasoning, and feature binding (e.g., Dessalegn & Landau, 2008, 2013; Loewenstein & Gentner, 2005; Miller, Patterson, & Simmering, 2016; Piccardi, Palermo, Bocchi, Guariglia, & D'Amico, 2015; Shusterman, Ah Lee, & Spelke, 2011). In correlational studies, children with larger spatial vocabularies outperform peers with more limited vocabularies on an equally broad range of spatial tasks (Ankowski, Thom, Sandhofer, & Blaisdell, 2012; Balcomb, Newcombe, & Ferrara, 2011; Gentner et al., 2013; Hermer-Vazquez, Spelke, & Katsnelson, 1999; Miller, Vlach, &

This article was published Online First March 26, 2020.

Marianella Casasola, Wendy S. Wei, Daniel D. Suh, Patricia Donskoy, and Ashley Ransom, Department of Human Development, Cornell University.

Daniel D. Suh is now at the Department of Applied Psychology, New York University. Wendy S. Wei is now at the Harvard Graduate School of Education.

We thank the children who participated and the preschool teachers and directors who generously hosted us. We are indebted to Zena Saunders as well as Angelica Gangemi, Tiffany Guo, Laura Sheridan, Kate Goldberg, Rachel Cohen, Terigray Nnanabu, Monet Bell, Lauren Ritter, Sarah Hudson, Veronica Palumbo, Allison Eatroff, Denise Ma, and the many other members of the Cornell Play and Learning Lab for their help with participant testing, transcription, and coding. Lisa M. Oakes, David H. Rakison, and Tamar Kushnir provided insightful feedback. This work was supported by an Evalyn Edwards Milman Fellowship through the Bronfenbrenner Center for Translation Research and an award from the Cornell Institute for Social Science. Portions of this work were reported at the 2013 and 2015 meetings of the Cognitive Development Society and the 2017 biennial meeting of the Society for Research for Child Development.

Correspondence concerning this article should be addressed to Marianella Casasola, Department of Human Development, Cornell University, 35 Beebe Hall, Ithaca, NY 14853. E-mail: mc272@cornell.edu

Simmering, 2017; Piccardi et al., 2015; Pruden et al., 2011; Sims & Gentner, 2008). These experimental findings, complemented by the well-documented association between children's spatial vocabulary and their spatial skills, motivate the present studies.

There is general consensus that spatial language can highlight spatial information that may otherwise go unnoticed (e.g., Dessalegn & Landau, 2013; Gentner, 2016; Miller et al., 2016; Shusterman et al., 2011). More controversial is the depth and duration by which spatial language supports spatial cognition. Whereas some argue that the facilitative effects of spatial language reflect enduring changes in spatial representations (Gentner, 2016; Gentner et al., 2013; Loewenstein & Gentner, 2005; Pyers et al., 2010), others counter that these effects are context-bound and can be attributed to priming effects that may quickly dissipate (e.g., Dessalegn & Landau, 2013; Papafragou, Hulbert, & Trueswell, 2008). Resolving this debate has been difficult because most experimental studies have taken a targeted approach to testing the effect of spatial language on spatial skills. With the goal of establishing how a particular spatial label (or set of labels) enhances performance on a specific spatial task, experimental studies typically present only task-relevant spatial language. Moreover, studies have rarely probed the persistence of gains from spatial language input. As exceptions, Dessalegn and Landau (2013) found that preschool children continued to benefit from spatial labels after a 4-s delay, and Loewenstein and Gentner (2005) reported maintenance after a 2-day delay. That is, spatial language can continue to impart its benefit for days, consistent with claims that spatial language strengthens spatial representations. One question is whether the facilitative effects of spatial language transfer to benefit spatial skills more broadly, beyond the targeted spatial skill and the immediate context in which the spatial language was provided.

Training studies of spatial skills, in which participants engage in activities designed to improve a particular spatial skill over several weeks, document benefits of the training on the trained spatial skill but also nontrained spatial skills (e.g., Casey, Erkut, Ceder, & Young, 2008; Hawes, Moss, Caswell, & Poliszczuk, 2015; Terlecki et al., 2008; Tzuriel & Egozi, 2010; Uttal et al., 2013; Vander Heyden, Huizinga, & Jolles, 2017; see Newcombe & Frick, 2010, for a review). Similar results have been reported in short-term training studies of word learning and early mathematical skills. In a 10-week training in which infants were taught labels for objects that shared a similar shape, Smith, Jones, Landau, Gershkoff-Stowe, and Samuelson (2002) noted that 17-month-old infants generalized this learning to novel object categories. They also demonstrated faster vocabulary growth for object labels compared with a control group of infants, documenting an effect of the training that extended to their acquisition of object labels beyond the context of the laboratory task. Likewise, Siegler and Ramani (2009) reported improvement in preschool's children's number line estimation and simple addition following several sessions of playing a linear board game. Might a parallel effect be evident with spatial language and spatial skills? That is, can enriching children's exposure to spatial language over a short-term training promote their spatial skills, generalizing benefits to spatial skills that may not easily align with the spatial language provided?

In support of this possibility, Pruden et al. (2011) noted that children's use of spatial language predicted stronger performance on several spatial tasks, even though there was not a clear link between the spatial language children used and the spatial tasks. Families with infants of 14 months were filmed once every four months until their children were 46 months. Caregivers who used more spatial language during this time, labeling the spatial features of objects (e.g., shape, size, dimension), not surprisingly, had children who also used more spatial language. These children, in turn, scored higher on spatial tasks when assessed at 4½ years compared with peers who had less exposure to spatial language and who produced less spatial language. Children's use of spatial language as toddlers predicted stronger spatial skills as preschoolers, suggesting that exposure to and use of spatial language benefits a broad array of spatial skills.

However, the naturalistic, longitudinal design of Pruden et al. (2011), ideal for capturing individual differences in caregiver and child spatial language use, cannot rule out the contributions of other variables. Children in their sample who used more spatial language may have done so because they engaged in more constructive play (e.g., blocks and puzzles). Children's engagement in constructive play has been positively associated with stronger spatial performance (e.g., Jirout & Newcombe, 2015; Öostermeijer, Boonen, & Jolles, 2014). Possibly, gains in spatial skill attributed to spatial language may have reflected the benefits of constructive play. In fact, using the same sample as Pruden et al., Levine, Ratliff, Huttenlocher, and Cannon (2012) found that toddlers who played with puzzles at 26 to 46 months scored higher on a spatial task at 4½ years. This result is not surprising given that children's exposure to spatial language and their engagement in constructive play often co-occur and may mutually support each other. Caregivers use more spatial language during constructive play relative to other contexts (Ferrara, Hirsh-Pasek, Newcombe, Golinkoff, & Lam, 2011). They also increase their use of spatial language when children tackle challenging jigsaw puzzles (Levine et al., 2012) or engage in structured tasks (Ferrara et al., 2011). It may be that spatial language and constructive play synergistically contribute to advances in children's spatial thinking and the positive effect of one may be bolstered by the presence of the other. Without an experimental approach, it remains unclear whether spatial language uniquely contributes to children's spatial thinking or may instead reflect the benefits of engaging in constructive spatial play.

The present two experiments were designed to test for a causal link between preschool children's exposure to spatial language and advances in their spatial skills. More specifically, we asked whether repeated experience to a broad array of spatial language in a naturalistic context, such as play, over several weeks transfers to benefit nontrained spatial skills. In the first experiment, we manipulated children's exposure to spatial language during constructive play and in a second experiment, their engagement in constructive spatial play to explore if spatial language promotes the development of spatial skills. In both experiments, children participated in five play sessions, each about 12 min in duration, over 3 to 4 weeks, approximating the quantity and spacing of input in other short-term training studies with preschool children (e.g., Siegler & Ramani, 2009).

The play activities were selected to provide a natural context for labeling a wide spectrum of spatial information, such as geometric shapes (e.g., "triangle," "square"), spatial relations (e.g., "left," "under"), spatial features and properties (e.g., "round," "edge"), orientation (e.g., "horizontal"), dimension (e.g., "large"), location (e.g., "corner"), and deictics (e.g., "there"). Critically, children's exposure to the spatial language only occurred during the play activities and never during the spatial tasks, which were given before the start of the play activities and one or more days after the play activities had ended. Nor did the play activities resemble or easily relate to the spatial tasks on which the children were assessed. This design was adopted as a stringent test of transfer of spatial language exposure during spatial play activities to non-trained spatial tasks and to the durability of any improvements.

We reasoned that exposure to a rich assortment of spatial language, across distinct play activities and over multiple sessions, would be more likely to promote generalization to nontrained spatial tasks than exposure to a narrow range of spatial terms in a single session (e.g., Vlach, 2014). If the primary mechanism by which spatial language supports spatial skills is through the temporary priming of attention, then there should be no evidence of transfer to nontrained spatial tasks in either experiment because children are only assessed on their spatial skills one or more days after the end of the training sessions. In contrast, if spatial language yields deeper processing of spatial features and promotes spatial representations, then facilitative effects should still be evident, even when children are assessed on their spatial skills several days after the training.

We tested 4-year-olds to match the participant age of previous studies examining the effect of spatial language on spatial skills (e.g., Dessalegn & Landau, 2008, 2013; Loewenstein & Gentner, 2005; Miller et al., 2016; Pruden et al., 2011). Further, this age has been argued to be within the ideal developmental period for promoting spatial thinking. By 4 years, motor and perspective-taking skills have sufficiently matured to allow children to benefit from exposure to more complex spatial concepts (Newcombe, 2010; Newcombe & Frick, 2010; Oudgenoeg-Paz, Leseman, & Volman, 2015).

The two spatial tasks in the present studies, the Children's Mental Transformation Task (CMTT, Levine, Huttenlocher, Taylor, & Langrock, 1999) and the Picture Rotation Task (PRT, Quaiser-Pohl, 2003), have been linked to children's spatial vocabulary, their engagement in constructive (i.e., spatial) play, or both. In particular, the CMTT has been linked to both children's spatial vocabulary and their engagement in constructive play (Levine et al., 2012; Pruden et al., 2011), while the PRT has been linked to children's engagement in constructive play (Öostermeijer et al., 2014). Therefore, these tasks were ideal for testing gains in spatial skills following our manipulation of spatial language exposure (Experiment 1) and the constructive nature of the spatial play activities (Experiment 2). In particular, these tasks shared minimal, if any, similarity to the play activities and were well suited for assessing transfer of the spatial language provided during the play activities to children's performance on these two spatial tasks.

Experiment 1

The goal in the first experiment was to test whether increasing children's exposure to spatial language results in advances in their spatial skills, holding constant their engagement in constructive play. At baseline, children's mental rotation, mental transformation, and spatial vocabulary were assessed. Once baseline measures were completed, children in each condition participated in the same adult-guided constructive play activities. There were five play sessions, each about 12 min, distributed over an average of four weeks. Children were randomly assigned to one of two conditions, a spatial-language or a play-only (control) condition. In the spatial-language condition, children heard generous amounts of spatial language during the activities, while those in the play-only (control) condition heard more general references to spatial features (see the Appendix). One or more days after the final play session, children were tested on posttest versions of the mental rotation and mental transformation tasks, as well as on the same Spatial Vocabulary Assessment administered at baseline. If spatial language promotes spatial cognition more than general references to these features (e.g., Gentner, 2003), and if such input generalizes to promote spatial skills more broadly, then children who hear a greater amount and diversity of spatial language (i.e., those in the spatial-language condition) should display stronger spatial skills at posttest than children who hear more general references to spatial information (i.e., those in the play-only condition). Alternatively, if children's engagement in constructive play suffices for promoting spatial skills (Casey et al., 2008; Jirout & Newcombe, 2015; Öostermeijer et al., 2014), then children in both groups should demonstrate equivalent gains in their spatial skills from baseline to posttest.

Method

Participants. The participants were 50 children of 4½ years (M = 4.75 years, SD = .37) with 22 girls and 28 boys recruited from local preschools. Children were fluent in English and were from predominantly middle-class, college-educated families. Children were White (n = 41), Asian or Asian American (n = 3), Black or African American (n = 2), biracial (n = 2), or of Hispanic (n = 2) descent. An a priori power analysis conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) based on the effect size reported for children's performance on the CMTT in Pruden et al. (2011) revealed that a sample of 23 children was required to achieve .80 power level with α set at .05.

An additional nine children were omitted from the final sample. Two children completed the baseline measures but were too young (they had just turned 3) and their participation was discontinued. Two male children did not complete the baseline measures because of shyness (n = 1) or inattentiveness (n = 1). Three female children (two in the play-only condition, one in the spatial-language condition) and two male children (spatial-language condition) completed the baseline measures but engaged in only one or two sessions of play. These five children did not want to interact with the experimenter for the play sessions, in several cases because the activities did not always include Lego sets. These five children did not differ from those in the final sample in age, $t(53) = 1.38$, $p = .17$, baseline performance on the CMTT, $t(52) = 1.13$, $p = .26$, the PRT, $t(53) = 1.52$, $p = .14$, Spatial Vocabulary Assessment, $t(52) < 1$, $p = .86$, and Peabody Picture Vocabulary Test (PPVT) general vocabulary, $t(51) < 1$, $p = .43$. Nor was there a significant difference in the distribution of gender, $\chi^2(1, N = 55) = .47$, $p = .49$, or condition, $\chi^2(1, N = 55) = .12$, $p = .73$, from the final sample.

Materials. Examples of the two spatial tasks, the CMTT and the PRT, and the Spatial Vocabulary Assessment are shown in Figures 1, 2, and 3, respectively. Children's percent correct was calculated for each task.

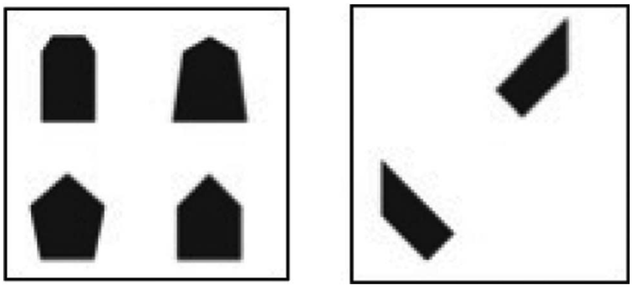

Figure 1. A sample item of the Children's Mental Transformation Task (Levine et al., 1999).

Children's Mental Transformation Task. Two distinct, abbreviated 10-item forms of the Children's Mental Transformation Task (CMTT), developed by Levine and colleagues (1999), were used. For each item, children were shown four shapes, arranged in a two-by-two matrix (see Figure 1). Beneath these shapes, children viewed two target pieces and were told, "We have two pieces here and four shapes here. Which one of these four shapes would these two pieces make if they were put together?" The target pieces could be rotated 0° or 45° with either horizontal or vertical translation. Unfortunately, the Cronbach's alphas for the first and second forms were below the level of acceptability, $\alpha = .57$ and 0.54, respectively. Results from this measure were not included in the analyses.

The Picture Rotation Task. The PRT is a mental rotation test for preschool children (Quaiser-Pohl, 2003). The original measure includes 16 items. For the purposes of the present study, an additional eight items were created, with an equal number at each of the four angles of rotation (see below). We then divided the task into two forms, each with 12 items. Each item displayed an upright target stimulus to the left of a vertical line. To the right of the line were three response options, two of which were mirror images of the target while a third option was identical to the target. The response options were rotated 45°, 90°, 135°, or 180° clockwise or counterclockwise, requiring children to mentally rotate the images to compare them to the target. Children's attention was first drawn to the target item and their left- or right-facing orientation was highlighted using a landmark in the room (e.g., "See this dog? This dog is looking at the window"). The experimenter then pointed to the three response options and asked the child to select the option that was exactly the same as the target (see Figure 2). Three

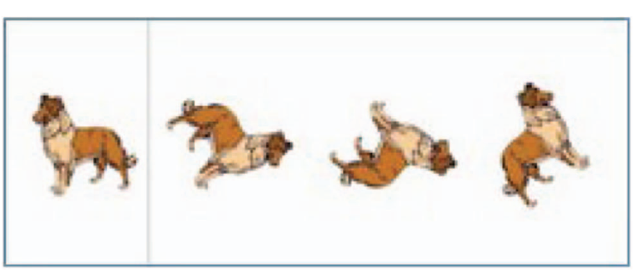

Figure 2. A sample item of the Picture Rotation Task, or the PRT (Quaiser-Pohl, 2003). See the online article for the color version of this figure.

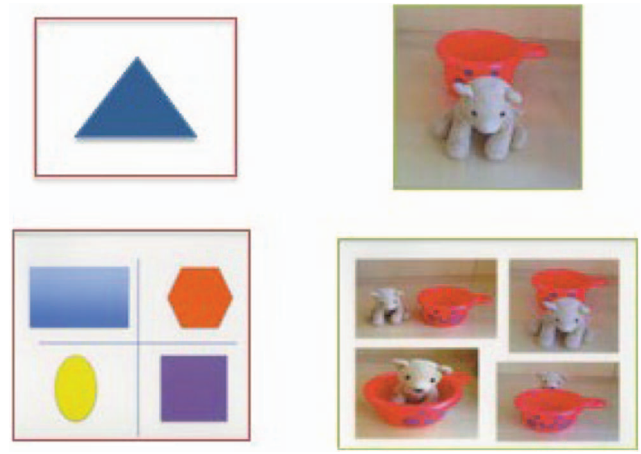

Figure 3. Sample items of the Spatial Vocabulary Assessment, with expressive items in the top row and receptive items on the bottom row. See the online article for the color version of this figure.

practice items were given with feedback, but no feedback was given on the 12 test items. Each form had acceptable internal consistency, Cronbach's $\alpha = .78$ and .74, respectively, with a Guttman's split half coefficient between the two sets of .80.

Spatial vocabulary assessment. Children's expressive and receptive vocabulary for geometric shapes and spatial configurations was assessed in a measure created for the present study. In the expressive task, children were shown 13 geometric shapes (circle, rectangle, square, triangle, oval, diamond, star, heart, rhombus, trapezoid, pentagon, hexagon, octagon) and eight spatial configurations of a stuffed bear in relation to a red bowl (behind, in, under, next to, in front of, away, middle/between, above) and were asked to name the shape ("Do you know the name of this shape?") or spatial relation ("Can you tell me where the bear is?"). To assess receptive language, children were asked to point to a shape or spatial relation from an array of two to four images. Children were assessed on their comprehension of four geometric shapes (diamond, semicircle, hexagon, pentagon), five spatial relations (below, under, away, behind, left), and two orientations (upside down, diagonal). The posttest measure had a high level of internal consistency with a Cronbach's alpha of .99. The baseline measure did not meet this same level of internal consistency with a Cronbach's alpha of .66, perhaps because several of the geometric shapes were unfamiliar to children.

General vocabulary. The Peabody Picture Vocabulary Test Version IV (PPVT-4; Dunn & Dunn, 2007) assessed children's receptive vocabulary. As a normed measure, this measure provided insight into how the present sample may compare to other children. It also was included to ensure that children in each condition were comparable in their receptive language skills.

Procedure. The protocol for the present study was approved by the university Institutional Review Board (Protocol #1210003363). Children were randomly assigned to the play-only or spatial-language condition. This randomization resulted in 24 children in the play-only condition (10 girls, 14 boys) and 26 children in the spatial-language condition (12 girls, 14 boys). Children were tested in a quiet corner or room of their preschool. Children sat at a table next to an experimenter. All sessions were filmed with a Canon FS200 camera.

Baseline measures. On the first day, children were given the CMTT followed by the PRT, with forms counterbalanced across children. They were then given the Spatial Vocabulary Assessment. On a second day, children completed the PPVT-IV.

Play activities. On a third day, children began the first of five play sessions which typically occurred once a week. The training took place over an average of four weeks. This timing did not differ across the two conditions, $F < 1$, $p = .49$. Each session included several play activities that were identical across the two conditions and were about 12 min in duration, with children engaging in a total of about 60 min of play by the end of all play sessions. Children created origami animals, scenes with magnetic shapes using a Djeco Geoforms set, small Lego structures, and played a shape-matching game. The first three authors conducted the majority of play sessions for children in both conditions and supervised the 10 additional experimenters who assisted with other play sessions. Children in each condition interacted with a comparable number of experimenters across the various play activities, $M = 3.42$, $SD = 1.51$ for children in the spatial-language condition and $M = 3.88$, $SD = 1.35$ for children in the play-only condition, $t(48) = 1.15$, $p = .41$.

Origami pig. On three distinct days, children created an origami pig face. The activity was taken from an online website (http://www.origami-instructions.com/easy-origami-piggy.html). Each step of the activity was printed in color, laminated, and bound into a small booklet as a visual reference for each step. Children selected a specific color of origami paper. They were asked to place the colorful side of their origami sheet face down and to orient it as a diamond. As the first step, children were instructed to take the top corner of their paper and fold it down to the bottom corner to create a horizontal crease through the center of the paper. They then opened the sheet and folded the left corner to the right corner to create a vertical crease. Children once again opened the sheet of paper. They then were guided to fold the right-hand corner to the center, at the intersection of the horizontal and vertical creases, and to do the same with the left corner of the paper. These folds created two small, equilateral triangles. They then took the top corner of their sheet and folded the paper in half, forming a pentagon. The top left and right corners of the pentagon were folded down to create ears. As the last step, children folded the bottom point of the pentagon up to create a mouth. Children took only the top sheet and folded it down to create the pig's mouth. As the last step, children were invited to draw eyes and the snout on their origami pig. This activity was about 4 min in duration.

Origami whale. On two separate days, children created a simple origami whale (http://www.origami-instructions.com/easy-origami-whale.html). Children selected a square sheet of origami paper, placed the colorful side face down, and oriented the paper as a diamond. Referencing an instruction manual with a color picture of each step, an experimenter instructed children to take the left corner of their sheet, place it on top of the right corner, and fold it down, creating a triangle. The child then was asked to open the triangle and to note the crease down the center of the paper that they had created. The child was then asked to take the left corner, place it in the center, and fold it down along the crease in the center, creating another triangle. The child did the same with the right-corner of the paper. The experimenter then asked children to fold down the top corner of the paper so that it touched the triangles. Then, children were asked to fold the paper in half along the middle crease. The paper was turned 90 degrees and the left-hand corner was folded to create the tail of the whale. The child was invited to draw the whale's eye. This activity was about 3 min in duration.

Magnetic board and shapes. On two different days, children created a scene from the Djeco Geoform Magnetic board set. The set included 43 magnetized, wooden geometric shapes (triangles, squares, rectangles, circles, semicircles, arcs) of varying sizes and color and two white magnetic boards joined by a hinge. Children chose a laminated card of a rocket ship, elephant, or horse. The card remained in view as the experimenter guided the child through creating the pattern using phrases such as "What do we need first?" or "Where does it go?" Only if needed, the experimenter guided the child in locating or indicating where to place a piece. This activity was about 5 min in duration.

Shape matching task. On four distinct days, children were asked to name and match four geometric shapes, a triangle, pentagon, hexagon, and octagon. The experimenter placed four, glittery red foam shapes in a 2×2 matrix and asked the child to name each one. The experimenter assured the child that it was fine if the child did not know a shape's name, noting that they sometimes had trouble naming some of these shapes themselves. The experimenter then placed two halves of a black foam shape, which was cut into two pieces along the vertical axis, below the four shapes. The pieces were placed so that the halves were rotated and separated by two inches. The experimenter kept a finger on each half so that children could not move the pieces. The children were asked which of the four shapes the halves would make if they were joined. Once children made their selection, the experimenter invited them to place the halves on top of their selection to see if it matched the shape. If there was not a match, the experimenter asked the child to try again and to select another shape. This game continued until children had the opportunity to match halves for each of the four shapes. This activity was about 3 min in duration.

Lego structures. On two distinct days, children built an age-appropriate Lego set with the experimenter. Children chose among the Creator Fire Rescue 6752 set in which children could build a fire truck, fire engine, or helicopter, the Lego Friends Stephanie's Pet Patrol 3935 set, in which children could build a truck and tracker and animal hutch, or the Lego Friends' Olivia's Speedboat 3837 set in which they could build a speedboat, sand castle, and beach mat with umbrella. The experimenter guided the child through each step of creating the structure, using the manual that was included with the set. Each Lego set took about 12 min to complete.

Experimenter language. All children engaged in the same play activities, but the amount and quality of experimenter spatial language differed across the two conditions. In the spatial-language condition, the experimenter labeled the orientation, placement, shape, features, and spatial configurations of the objects during the play activities. In the play-only (control) condition, the experimenter used deictic or general terms (e.g., "shape," "there") rather than the more semantically rich spatial words (e.g., "triangle," "under"). The Appendix provides examples of the language input in each condition and for each play activity.

To create naturalistic interactions, the experimenter did not follow a verbatim script but rather a general one for each play activity. The first three authors, who conducted or supervised the play activities for all participants, agreed on the specific spatial words to provide for each activity in the spatial-language condition and which general language words to use in the play-only condition. Care was taken to ensure that attention was drawn to the same spatial features and that the amount of adult interaction and scaffolding was equated across the two conditions. Other experimenters were trained by the first three authors and were required to observe each condition prior to then conducting a play activity under the supervision of one of the authors.

To initially ensure adherence, five observers, unaware of child condition, viewed the videos of the magnet board task, the first session for all participants except for one child who was missing the video of this session. For this child, the second magnet-board session was used. The observers viewed the videos and were asked to guess the condition based on experimenter language. This task was an especially stringent test of experimenter language in each condition because the task relied heavily on referring to geometric shapes. Each video was viewed by one observer. These observers accurately predicted the correct condition for all but five children ($n = 45$), yielding a Cohen's Kappa of .92. For three videos, observers incorrectly guessed that the child was in the spatial-language condition, and for two videos they incorrectly guessed that they were in the play-only condition. This initial evaluation of the language input documented that experimenters were abiding by the spatial language guidelines for each condition.

Using the Children Language Analysis program, CLAN (MacWhinney, 2000), videos of all play activities were transcribed with the exception of the Lego structure, which initially was not video-recorded because of its longer duration. The number of experimenter and child tokens and types was calculated for each play activity, generated from the CLAN software. Tokens refer to the total number of words, including repetitions of the same word. Types refer to the number of unique words and documents the diversity of words. The utterance "Place the blue triangle next to the red triangle" has nine tokens and seven types. Tokens capture the amount of language input, whereas types reflect one aspect of language quality (e.g., Rowe, 2012).

Next, experimenter and child tokens and types were categorized as either spatial or nonspatial using the coding scheme developed by Cannon, Levine, and Huttenlocher (2007). Spatial words were those that referred to spatial dimensions (e.g., "big," "long," "wide"), geometric shapes (e.g., "circle," "square," "hexagon"), spatial features and properties (e.g., "round," "corner," "flat"), orientations and transformations (e.g., "upside down," "turn," "flip"), location and directions (e.g., "under," "left," "near"), continuous amount (e.g., "whole," "half," "a lot"), pattern words (e.g., "next," "first" "last"), and deictics (e.g., "here," "there," "where"). A second observer transcribed 20% of the play activities. There was high interrater reliability across transcribers in the number of types and tokens calculated with intraclass correlations of $r = .995$ and .998, respectively. Any discrepancies between the transcripts were resolved by the fifth author, but these were typically attributable to spelling differences.

As a second manipulation check, the proportion of experimenter spatial language provided to children was compared across conditions. Seven children ($n = 6$ spatial-language condition) were missing a video from one play activity (e.g., one of the three origami pigs), and two children in the play-only condition were missing videos from two activities. Because 11 of the total 550 videos were missing (2% of all videos), we averaged the amount of spatial and nonspatial language tokens (i.e., number of words) and types (i.e., unique words) across all activities for each child, and then calculated the proportion of this input that was spatial language. We included the average number of nonspatial types or tokens (calculated as total types or tokens minus spatial types or tokens) and the average time spent in the play activities as covariates in many of the analyses to ensure that any effects of spatial language were not due to differences in overall language input or duration of the play sessions.

An analysis of covariance (ANCOVA), with the average number of nonspatial experimenter tokens and the average duration of the play sessions included as covariates, yielded only a significant effect of condition on the proportion of experimenter spatial tokens, $F(1, 46) = 44.33$, $p < .001$, $\eta_p^2 = .49$. Children in the spatial language condition heard a higher proportion of spatial language tokens ($M = .17$, $SD = .02$) than children in the play-only control condition ($M = .12$, $SD = .01$). A similar analysis with the proportion of experimenter spatial types also yielded only a significant effect of condition, $F(1, 46) = 68.47$, $p < .001$, $\eta_p^2 = .60$. Experimenters provided a significantly higher proportion of spatial types in the spatial-language ($M = .22$, $SD = .02$) than play-only ($M = .16$, $SD = .02$) condition. In sum, experimenters provided significantly more (i.e., tokens) and a greater variety (i.e., types) of spatial language to children in the spatial-language than those in the play-only condition.

Posttest measures. Posttest measures were given one to 14 days after the last play session. Children were given alternate forms of the CMTT and PRT at posttest but the same version of the Spatial Vocabulary Assessment. Children in the play-only condition did not differ in the number of days between the last play session and the first day of posttests ($M = 4.67$ days, $SD = 3.23$ days) compared with those in the spatial-language condition ($M = 4.88$ days, $SD = 3.99$ days), $t = -.21$, $p = .83$.

Children's responses on each task were coded on paper at the time of testing. Two coders, unaware of child condition or study hypotheses, used the video of the child sessions to independently score all participants' responses on each task. There were no discrepancies in coding by the video coders. After completing the posttest measures for the present study, children completed three other measures on a separate day. The results of these tasks are not reported because they were not given at baseline and are the focus of a distinct research question.

Results

Descriptive statistics for each measure are listed in Table 1, and the correlation among the measures are listed in Table 2. Children performed above chance on each measure at baseline and posttest. There were no significant effects of condition at baseline (Spatial Vocabulary, PPVT-IV, $ps > .46$; PRT, $p > .14$) and no significant effect of child sex at baseline or posttest, all $ps > .22$. We subsequently collapsed child sex in the remaining analyses.

Relations between measures. As can be seen in Table 2, with the exception of the PPVT, which was normed for age, each measure was significantly correlated with child age at both baseline and posttest. We thus conducted partial correlations to control for the effect of child age. In this partial correlation,

Table 1
For Each Condition of Experiment 1, the Mean (and Standard Deviation) of Children's Percent Correct on Each Measure

Condition	Spatial-language condition		Play-only control condition	
	Baseline	Posttest	Baseline	Posttest
PRT				
M (*SD*)	.446 (.227)**	.637 (.250)**	.538 (.210)**	.510 (.264)**
95% CI	[.354, .537]	[.534, .740]	[.450, .627]	[.399, .622]
Spatial vocabulary				
M (*SD*)	.685 (.111)**	.773 (.098)**	.707 (.094)**	.750 (.079)**
95% CI	[.639, .731]	[.733, .813]	[.667, .747]	[.716, .783]
PPVT (single administration)				
M (*SD*)	84.1 (21.2)**		79.4 (24.5)**	
95% CI	[75.5, 92.7]		[69.0, 89.7]	

Note. PRT = Picture Rotation Task; PPVT = Peabody Picture Vocabulary Test.
** $p < .01$.

there was no significant relation across measures at baseline or at posttest.

Condition differences across measures? We conducted analyses of variance to test for condition differences in children's general vocabulary, spatial vocabulary, and, critical to the goals of the study, their mental rotation. To best address the central question of our study, we conducted a series of regression analyses with the spatial task, in which we assessed whether the proportion of experimenter spatial language during the play activities accounted for a significant amount of the variance in children's posttest mental rotation performance. In all analyses, we used the proportion of experimenter spatial language as well as the proportion of child spatial language in analyses that examined spatial language tokens (or in other analyses, those that examined spatial language types) because the proportion of spatial language, whether for the experimenter or for the child, had lower multicollinearity in the regression analyses than the counts of spatial tokens and types, which were significantly correlated with the number of nonspatial tokens and types.

General vocabulary. There were no condition differences in children's general vocabulary, assessed with the PPVT-IV, $F < 1$, $p = .47$, indicating that children in each condition had comparable receptive language skills.

Table 2
Correlations Between Child Age, the Baseline and Posttest Measures, and General Vocabulary in Experiment 1

Age, task	1	2	3	4	5	6
1. Child age		.352*	.360*	−.112	.418*	.342*
2. BL PRT			.237	.001	.576*	.092
3. BL spatial vocabulary		.107		.132	.330*	.489*
4. General vocabulary		.045	.201		−.039	.082
5. PT PRT		.504*	.169	−.004		.240
6. PT spatial vocabulary		−.035	.418*	.140	.113	

Note. The correlations between child age, the baseline and posttest measures, and general vocabulary are listed in the top half of the table, above the diagonal. Partial correlations that control for child age appear below the diagonal. BL PRT = baseline PRT; BL = baseline; PT PRT = posttest PRT; PT = posttest.
* $p < .05$.

Spatial vocabulary. One girl in the spatial-language condition was more than three standard deviations below the mean for all children at posttest and was omitted from the analysis. Accounting for baseline spatial vocabulary, the average number of experimenter nonspatial tokens, and the average time children engaged in play activities, a one-way analysis of covariance (ANCOVA), examining the effect of condition on children's spatial vocabulary at posttest, yielded only a significant effect of baseline spatial vocabulary, $F(1, 44) = 30.78$, $p < .001$, $\eta_p^2 = .41$. There was no effect of condition, $F(1, 44) = 2.23$, $p = .14$. Both children in the spatial-language condition and those in the play-only condition demonstrated significant gains in spatial vocabulary from baseline to posttest, paired $t(24) = 7.00$, $p < .001$ and paired $t(23) = 2.16$, $p = .04$, respectively.

Mental rotation. Did children in the spatial language condition demonstrate stronger performance on the mental rotation task at posttest than those in the play-only condition? A one-way analysis of covariance (ANCOVA) examined condition effects on children's percent correct at posttest, with baseline performance, the average number of nonspatial experimenter tokens and average time in play activities as covariates. The analysis yielded a significant effect of baseline performance, $F(1, 45) = 32.97$, $p < .001$, $\eta_p^2 = .42$, and condition, $F(1, 45) = 9.43$, $p = .004$, $\eta_p^2 = .17$. Children in the spatial-language condition improved from baseline to posttest, paired $t(25) = 3.96$, $p = .001$, but those in the play-only condition did not, paired $t < 1$, $p = .34$ (see the left-hand graphs of Figure 4). There were no other significant effects.

Does the proportion of experimenter spatial tokens account for children's posttest mental rotation performance? We conducted a series of multiple linear regressions to document if the amount and diversity of experimenter spatial language during the play sessions accounted for a significant amount of the variance in children's posttest performance on the PRT mental rotation task. In the first model, baseline PRT performance, child age, child general vocabulary, the average duration of the play activities, and the average number of experimenter and child nonspatial tokens were included. This model was significant, $F(6, 43) = 5.72$, $p < .001$, with an adjusted R^2 of .37. Children's baseline score, $\beta = .50$, $t = 4.08$, $p < .001$, and age,

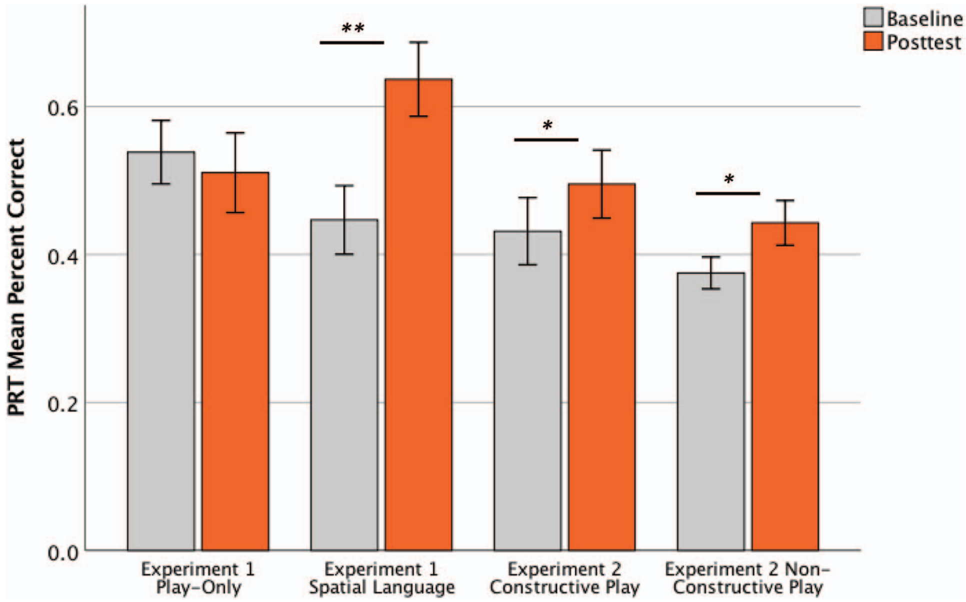

Figure 4. Children's mean percent correct on the Picture Rotation Task (PRT) at baseline (gray or white bars) and posttest (orange or dark bars) in each condition of Experiment 1 (the two graphs on the left) and Experiment 2 (the two graphs on the right). Errors bars represent one standard error. Asterisks indicate a significant change from baseline to posttest. * $p < .05$. ** $p < .01$. See the online article for the color version of this figure.

$\beta = .33$, $t = 2.51$, $p = .02$, were the only significant predictors of the PRT posttest score. A second model that added condition was significant, $\Delta F(1, 42) = 5.07$, $p = .03$, with a ΔR^2 of .05 and an adjusted R^2 of .42, indicating that condition, $\beta = .36$, $t = 2.25$, $p = .03$, explained a significant amount of the variance in children's posttest PRT performance beyond child age and baseline scores.

In the next set of analyses, shown in Table 3, we replaced condition with the proportion of experimenter spatial tokens. We omitted variables that were not significant in the previous models (i.e., child general vocabulary, the average duration of the play activities) with the exception of the average number of experimenter and child nonspatial tokens. As can be seen in Table 3, baseline PRT performance, child age and the average number of experimenter and child nonspatial tokens significantly predicted posttest PRT performance, $F(4, 45) = 8.72$, $p < .001$, with an adjusted R^2 of .39 (Model 1). Children's baseline PRT scores, $\beta = .50$, $t = 4.11$, $p < .001$, age, $\beta = .30$, $t = 2.46$, $p = .02$, and the average number of experimenter nonspatial tokens, $\beta = .25$, $t = 2.06$, $p = .046$, significantly predicted posttest scores. In Model 2, the proportion of experimenter spatial tokens was also included and significantly predicted posttest PRT performance, $\Delta F(1, 44) = 6.59$, $p = .01$, with a ΔR^2 of .07 and an adjusted R^2 of .45. In Model 2, children's baseline PRT, $\beta = .62$, $t = 4.72$, $p < .001$, age, $\beta = .26$, $t = 2.18$, $p = .04$, and the proportion of experimenter's spatial tokens, $\beta = .35$, $t = 2.57$, $p = .01$, accounted for a significant amount of the variance in children's posttest PRT scores. The average number of experimenter nonspatial tokens was no longer significant once the proportion of experimenter

Table 3
Summary of Hierarchical Regression Analysis for Variables Predicting Children's Posttest PRT Scores (N = 50)

	Model 1			Model 2			Model 3		
Variable	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Age	.017	.007	.302*	.015	.007	.255*	.012	.007	.218
Baseline PRT	.581	.141	.497*	.703	.141	.601*	.707	.142	.605*
Ave. exptr. nonspatial tokens	.001	.000	.247*	.000	.000	.039	.000	.000	.035
Ave. child nonspatial tokens	.000	.001	−.019	.001	.001	.056	.001	.001	.111
Prop. exptr. spatial tokens				2.808	1.094	.347*	.235	1.200	.291
Prop. child spatial tokens							.464	.499	.132
R^2		.437			.510			.520	
F for Δ in R^2		8.72*			6.59*			.866	

Note. Exptr. = experimenter; Ave. = average; Prop. = proportion; PRT = Picture Rotation Task.
* $p < .05$.

spatial tokens was added to the model. A third model that added the proportion of children's spatial tokens was not significant, $\Delta F(1, 43) = .25$, $p = .62$. The results of this regression analysis document that the quantity of experimenter's spatial language significantly predicted children's posttest mental rotation scores.

Does the proportion of experimenter spatial types account for children's posttest mental rotation performance? We also examined whether the proportion of experimenter spatial types, that is, unique spatial words, accounted for a significant amount of the variability in children's posttest PRT scores (see Table 4). Model 1 with children's baseline PRT, age, and the average number of experimenter and child nonspatial types was significant, $F(4, 45) = 8.16$, $p < .001$, with an adjusted R^2 of .37. Children's baseline PRT, $\beta = .49$, $t = 4.03$, $p < .001$, and age, $\beta = .27$, $t = 2.20$, $p = .03$, predicted PRT posttest scores. Model 2 in Table 4 added the proportion of experimenter spatial types and yielded a significant model, $\Delta F(1, 44) = 7.43$, $p = .009$, with a R^2 change of .08 and an adjusted R^2 of .45. Children's baseline PRT, $\beta = .59$, $t = 4.96$, $p < .001$, age, $\beta = .25$, $t = 2.20$, $p = .03$, and the proportion of experimenter spatial types, $\beta = .34$, $t = 2.72$, $p = .009$, predicted a significant amount of the variance in the posttest scores. A third model with the average number of children's spatial types was not significant, $\Delta F(1, 43) = 1.49$, $p = .23$.

The results of the regression analyses show that the diversity of the experimenter's spatial language significantly predicted children's posttest mental rotation scores. In sum, the quantity and quality of children's spatial language exposure partially accounted for posttest scores on the mental rotation task.

Exploratory analyses. As a first step in understanding how exposure to spatial language promoted children's performance on the mental rotation task, we conducted a set of exploratory analyses. In particular, we explored whether the effect of spatial language exposure was greater for the more difficult items, those rotated at higher angles of rotation (e.g., Berneiser, Jahn, Grothe, & Lotze, 2018; Shepard & Metzler, 1971; Wright et al., 2008). To this end, we compared the effect of condition on children's posttest performance on task items with the lower angles of rotation (e.g., 45 and 90 degrees, $n = 6$ items) versus higher angles of rotation (e.g., 135 and 180 degrees, $n = 6$ items). A repeated-measures analysis of covariance (ANCOVA) with angle of rotation (lower vs. higher) as the dependent variables and baseline PRT performance and the average number of experimenter nonspatial tokens as covariates yielded a significant condition x rotation angle interaction, $F(1, 44) = 12.18$, $p = .001$, $\eta_p^2 = .22$. There was no effect of condition for children's posttest PRT performance at the two lower angles of rotation, $F < 1$, $p = .63$, but a significant one for their posttest performance at the two higher angles of rotation, $F(1, 46) = 16.56$, $p < .001$, $\eta_p^2 = .27$.

Did the degree to which the proportion of experimenter spatial language predict posttest performance also differ across PRT task items with lower versus higher angles of rotation? We conducted the same set of regression models examining children's performance on the items with lower angles of rotation (45 and 90 degrees) separately from those with higher angles of rotation (135 and 180 degrees).

Lower angles of rotation: Spatial tokens. For the PRT items in which the correct response was rotated 45 or 90 degrees, a linear regression with baseline PRT performance for each of these six items, child age and the average number of experimenter and child nonspatial tokens was significant, $F(4, 45) = 6.99$, $p < .001$, with an adjusted R^2 of .33. Adding the proportion of experimenter spatial tokens failed to account for additional variance in a second model, $\Delta F(1, 44) < 1$, $p > .95$, but a third model with the proportion of children's spatial tokens was significant, $\Delta F(1, 43) = 5.00$, $p = .03$, with an adjusted R^2 of .37. Children's baseline PRT, $\beta = .58$, $t = 4.73$, $p < .001$, and the proportion of child spatial tokens, $\beta = .34$, $t = 2.24$, $p = .003$, predicted a significant amount of the variance on children's posttest performance for items rotated 45 and 90 degrees.

Lower angles of rotation: Spatial types. A regression model with baseline PRT performance for the two lower angles of rotation, child age, and the average number of experimenter and child nonspatial types was significant, $F(4, 45) = 6.76$, $p < .001$, with an adjusted R^2 of .32. A second regression model that added the proportion of experimenter spatial types was not significant, $\Delta F(1, 44) < 1$, $p = .86$, whereas a third model that added the proportion of children's spatial types was significant, $\Delta F(1, 43) = 5.42$, $p = .03$ with an adjusted R^2 of .37. Children's baseline PRT for the lower angles of rotation, $\beta = .53$, $t = 4.41$, $p < .001$, and the proportion of child spatial types, $\beta = .39$, $t = 2.33$, $p = .03$, predicted a significant amount of the variance on children's performance for posttest items rotated at 45 and 90 degrees.

These results suggest that it was children's, and not experimenters', use of spatial language that accounted for the posttest performance on the mental rotation task for items rotated at 45 and 90

Table 4
In Experiment 1, the Summary of Hierarchical Regression Analysis for Variables Predicting Children's Posttest PRT Scores (N = 50)

	Model 1			Model 2			Model 3		
Variable	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Age	.015	.007	.268*	.014	.007	.251*	.012	.007	.204
Baseline PRT	.574	.142	.490*	.695	.140	.594*	.678	.140	.580*
Ave. exptr. nonspatial types	.004	.002	.212	.001	.003	.051	.001	.003	.062
Ave. child nonspatial types	−.001	.002	−.052	.001	.002	.053	.002	.002	.114
Prop. exptr. spatial types				2.399	.880	.341*	1.579	1.103	.224
Prop. child spatial types							.778	.637	.192
R^2		.421			.084			.017	
F for Δ in R^2		8.164*			7.43*			1.49	

Note. Exptr. = experimenter; Ave. = average; Prop. = proportion; PRT = Picture Rotation Task.
* $p < .05$.

degrees, the two lower angles of rotation on the mental rotation task.

Higher angles of rotation: Spatial tokens. For the higher angles of rotation, those rotated 135 and 180 degrees, a linear regression with baseline PRT performance for these six items, child age and the average number of experimenter and child nonspatial tokens was significant, $F(4, 45) = 5.81$, $p = .001$, with an adjusted R^2 of .28. A second regression model that added the proportion of experimenter spatial tokens was also significant, $\Delta F(1, 44) = 10.60$, with an adjusted R^2 of .41. Children's baseline PRT on these items with higher angles of rotation, $\beta = .41$, $t = 3.46$, $p = .001$, age, $\beta = .33$, $t = 2.71$, $p = .006$, and the proportion of experimenter spatial tokens, $\beta = .44$, $t = 3.26$, $p = .002$, predicted a significant amount of the variance on the posttest items rotated at 135 and 180 degrees. A third model that added the proportion of children's spatial tokens was not significant, $\Delta F(1, 43) < 1$, $p = .91$ with an adjusted R^2 of .37.

Higher angles of rotation: Spatial types. A regression model with baseline PRT performance on the items rotated 135 and 180 degrees, child age, and the average number of experimenter and child nonspatial types was significant, $F(4, 45) = 5.48$, $p = .001$, with an adjusted R^2 of .27. A second regression model that added the proportion of experimenter spatial types also was significant, $\Delta F(1, 44) = 13.22$, $p = .001$, with an adjusted R^2 of .42. In this second model, children's baseline PRT for the items with the two higher angles of rotation, $\beta = .41$, $t = 3.54$, $p = .001$, child age, $\beta = .34$, $t = 2.91$, $p = .006$, and the proportion of experimenter spatial types, $\beta = .44$, $t = 3.64$, $p = .001$, predicted a significant amount of the variance on the posttest items rotated at 135 and 180 degrees. Finally, a third model that added the proportion of children's spatial types was not significant, $\Delta F(1, 43) < 1$, $p = .96$.

For the higher angles of rotation, experimenter spatial language was related to children's posttest performance, with the amount and quality of experimenter spatial language accounting for a significant amount of variability on children's performance for the more challenging items of the mental rotation task. These results show that children's exposure to spatial language was especially effective in supporting children's posttest performance as task difficulty increased.

Discussion

Children who heard a greater amount and diversity of spatial language during the play sessions demonstrated significantly greater gains in their mental rotation relative to peers who heard more general references during the play sessions. This result cannot be attributed to differences in the amount of nonspatial language input nor in the duration of the play activities. Neither of these variables was related to children's performance on the spatial task at posttest. Rather, as the amount and diversity of spatial language increased, so did children's performance on the mental rotation task at posttest and seemed to do so for items at higher angles of rotation.

The present results are the first to document a causal link between children's exposure to spatial language during naturalistic play interactions and significant improvements in their mental rotation. The results are noteworthy in documenting a transfer and maintenance effect of spatial language input on children's mental rotation. The present result is especially striking given that neither the spatial language nor the play activities were tailored to promote the mental rotation task, nor was spatial language ever used during the spatial tasks. Rather, children's experience with spatial language during the play activities transferred to promote their mental rotation. In addition, the benefits of the spatial language exposure were not fleeting. The facilitative effects remained evident a day or more after children had completed the play sessions. These findings provide experimental evidence for Pruden et al.'s (2011) assertion that children's experience with spatial language promotes aspects of their spatial thinking.

We conducted a second experiment to further examine how experience with spatial language during play interactions promotes children's mental rotation. We were particularly interested in documenting whether spatial language exerted its facilitative effect in this first experiment by amplifying the effects of constructive play (e.g., Lego builds, origami, creating a pattern). If so, then gains in spatial thinking should only be evident when children's exposure to the spatial language is in the context of constructive play and be mitigated (or perhaps absent) in the context of nonconstructive play. Alternatively, if spatial language uniquely contributes to the development of children's mental rotation, independently of a constructive element of spatial play, then a facilitative effect of spatial language should be evident across both constructive and nonconstructive play activities. We tested between these two possibilities in a second experiment.

Experiment 2

The goal of Experiment 2 was to compare the effect of spatial language exposure on children's spatial skills across constructive and nonconstructive play activities (i.e., those lacking a building or constructive element). Children in Experiment 2 were randomly assigned to one of two spatial play conditions. Those in the constructive-play condition participated in play activities with a constructive element, such as assembling blocks or creating patterns, similar to the constructive play in Experiment 1. Those in the nonconstructive-play condition participated in activities that, although containing spatial information, lacked a constructive element. Examples of the nonconstructive play activities were a shape scavenger hunt, playing with geometric shapes on a board without creating a pattern, searching for shapes, orientations, and spatial configurations in a picture book, pasting geometric paper shapes to make a collage, and completing a connect-the-dots drawing. Importantly, activities in each condition were designed to provide comparable amounts and diversity of spatial language and exposure to spatial information. If exposure to spatial language in the presence of spatial information suffices for promoting children's spatial skills, then children in both conditions should demonstrate equivalent gains in spatial skills. Alternatively, if constructive play boosts the facilitative effects of spatial language on children's spatial skills, then children in the constructive play condition should demonstrate greater gains in their spatial skills than children in the nonconstructive-play condition.

This second experiment also sought to replicate the results of the first experiment with a different sample of participants. We recruited children from Head Start centers to explore whether our findings generalize to a sample of 4-year-olds whose linguistic and spatial skills at baseline might differ from those of our predominantly middle-income sample in Experiment 1 (e.g., Hart & Risley, 1995; Jirout & Newcombe, 2015; Levine et al., 1999, 2005; Whitehurst et al., 1994). Finally, to examine the effect of spatial language exposure on children's performance on the mental transformation task, we increased the number of items in the mental transformation task to improve the scale reliability of the former abbreviated versions of this task.

Method

Participants. The final sample was 34 children of 4 ½ years ($M = 4.62$ years, $SD = 0.51$) with 15 boys and 19 girls from two Head Start centers. The centers served children from low-income families. All children were fluent in English. Children were White ($n = 10$), Black or African American ($n = 13$), or of Hispanic ($n = 11$) descent. Two additional girls did not complete the study because they lost interest in participating in the play activities. Sample size was determined by conducting a post hoc power analysis with G*Power (Faul et al., 2007) using the regression analysis of the proportion of experimenter spatial tokens on children's posttest mental rotation. This power analysis indicated that we achieved a power level of .999, with our sample size in Experiment 1, and an observed effect size of Cohen's $f^2 = .82$. With this effect size, an a priori power analysis revealed that a sample of 14 children was required for Experiment 2 with power ($1 - \beta$) set at .80 and α set at .05.

Materials. We used the PRT and the PPVT-IV from Experiment 1. We revised the CMTT for this second experiment. In Experiment 1, similar to Pruden et al. (2011) and Levine et al. (2012), we used an abbreviated version of the task with 10 items. For Experiment 2, we increased the abbreviated version of the CMTT to 16 items and created two forms, counterbalanced for use at baseline versus posttest. Each form had an equal number of items ($n = 4$) with vertical translation, lateral translation, lateral translation with 45-degree rotation, and vertical translation with 45-degree rotation. One form of the CMTT had a Cronbach's alpha of .58 while the other form had a Cronbach's alpha of .24, indicating a scale reliability lower than needed for acceptability. Results from this measure were not included in the analyses.

For the Spatial Vocabulary Assessment, we also modified this task and asked children to identify an additional six shapes (rectangle, octagon, oval, triangle, circle, and square) to sample a broader array of children's receptive vocabulary for geometric shapes. The Cronbach's alpha for the baseline measure was .86 and .83 for the posttest measure, indicating acceptable levels of internal consistency. Children's responses on each task were scored on paper and later, independently scored via video by two other coders, unaware of condition. There were no discrepancies between the two video coders. Children's percent correct was calculated on all tasks.

Procedure. The university Institutional Review Board (Protocol #1210003363) approved the present protocol. Children were randomly assigned to the constructive-play or nonconstructive-play condition. This randomization resulted in 17 children in the constructive play condition (nine girls, eight boys), and 17 children in the nonconstructive play condition (10 girls, seven boys). All testing and play sessions were video recorded. A team of four experimenters tested children in both conditions of Experiment 2.

Baseline measures. On the first day, children were given the CMTT, followed immediately by the PRT, then the Spatial Vocabulary assessment. These baseline measures were given in a single session, with breaks between tasks provided if needed.

Training: Play activities. We modified the timing of the play sessions to fit the academic schedule of the preschool centers where testing took place. The play sessions were given twice a week across two and a half weeks. Children engaged in one to two play activities for each play session. The play sessions varied in duration from 7 to 15 min, depending on the specific play activities for the session. Children in the two conditions did not differ in the number of days it took to complete the five play sessions, $t = 1.47$, $p = .15$.

Similar to Experiment 1, the experimenter did not follow a verbatim script when interacting with children in each condition. Instead, they used general guidelines to ensure that they labeled a variety of spatial information relevant to the play activities, such as geometric shapes, spatial configurations and locations, dimensions, orientations, amounts, and patterns. Guidelines were created for each play activity, with a list of spatial words to be targeted. For many activities, they asked children to name geometric shapes to encourage spatial language use. If a child could not name a shape, the experimenter invited the child to count the number of sides of that shape and, if the child could still not generate the shape name, labeled it for the child, and asked the child to repeat the shape name.

Constructive play activities. The constructive play activities in Experiment 2 were similar to those in Experiment 1. Children created the same origami pig and origami whale used in Experiment 1. They also created a pattern with magnetic shapes using the MindWare Imagination Magnets board set, choosing among three laminated cards, a tower, giraffe, or elephant. Children also participated in a number of constructive play activities that were created for the purposes of the second study. These are described in more detail below.

Block-building. In this task, children were shown a laminated card with a shape and asked to name the shape pictured on the card. They were then given a pair of foam blocks, asked to name those shapes, and invited to explore how to combine them to match the geometric shape on the card. As one example, children were shown a picture of a green square and were given two right-angle triangles to join into the square in the picture. The pictured shape was sized to match the foam blocks so that children could arrange the blocks on top of the image and determine if their block configuration matched the picture. If children had difficulty with the task, the experimenter would offer suggestions on how to orient the foam blocks. Children completed combining pairs of foam shapes for 10 laminated cards, an activity that was about 10 min in duration.

Magnatiles. In this activity, children were given a geometric shape cut out of a foam sheet and asked to name the shape. Children were then shown a magnatile and asked to name its shape. The child was given a second magnatile, asked to name it, and then asked to combine the magnatiles to match the foam shape. The foam shapes were sized to the magnatiles so that children could place the magnatiles on top of the foam shape and determine if they had succeeded in creating the target shape. Children used magnatiles to match foam shapes that were a rectangle (two square magnatiles), a blue foam square (four equilateral triangles), a large foam orange triangle (four equilateral triangles), a pink trapezoid (three isosceles triangles) and a green hexagon (six equilateral triangles). This activity was about 7 min in duration.

Lego structures. Children were guided in building a Lego alligator or a Lego flower on one day of play activities, and on another day of play, a Lego ghost or a tiger, all from the Lego Creator set. The experimenter used a laminated booklet that outlined each step of creating the structures, downloaded from the monthly builds from the Lego website, and provided references to the shape of the pieces and the spatial configuration or locations for placing the pieces onto the structure. The duration for creating the alligator or tiger was about 15 min and for creating the flower or ghost was about 10 min.

Jigsaw puzzles. Children assembled a 35-piece jigsaw puzzle. Children chose between a puzzle that pictured Thomas the Train or one with a unicorn in a meadow. Once the child selected which puzzle to build, the experimenter helped the child overturn each piece of the puzzle so that the image was visible. The experimenter then placed the box cover in front of the child for reference and asked the child to indicate which section they wanted to build first. Once the child pointed to a section, the experimenter invited the child to find those pieces. The experimenter provided scaffolding when needed. This activity was about 15 min in duration.

Nonconstructive play activities. The nonconstructive-play activities were similar to the constructive-play activities but lacked a constructive component, although they often provided exposure to geometric shapes and other spatial features. The activities were designed to provide an equivalent amount and variety of spatial language as children in the constructive-play condition. Children completed connect-the-dot animal drawings, created a collage using paper circles, squares, triangles and rectangles, searched for shapes and spatial configurations in an I Spy book and in a scavenger hunt at their preschool, played a shape memory game, manipulated magnetic shapes with the Imagination Magnet set without the laminated cards, and sorted foam shapes with the Lauri Shape and Color Sorter toy. These activities allowed the experimenter to label similar spatial information as the activities in the constructive-play condition. Each activity is described in more detail below.

Connect-the-dots. In this nonconstructive play activity, children were guided in connecting numbered or lettered dots on a sheet to create distinct geometric shapes that formed a pig's face. The experimenter directed the child to the first dot, indicated with the number one, and guided the child to draw a line to the next dot in the sequence until they had created a large circle, the outline of the pig's face, with two triangles, the pig's ears. Next, inside the circle, the child was guided to connect five dots to create a pentagon and then to do the same with another set of five dots, to create another pentagon. Each pentagon was the pig's eyes. The child then was asked to connect the dots in the center, which created a hexagon, the pig's snout, and within that hexagon, two squares. As they created each geometric shape, the experimenter would ask the child to name the shape. In this way, children were given exposure to the shape names of circles, square, triangle, pentagon and hexagon as well as their placement on the sheet. This task was about 6 min in duration.

Shape collage. Children were given geometric paper shapes of distinct colors to paste onto a large sheet of construction paper. Children were given six shapes, selected at random from the following shapes: a circle, heart, star, triangle, square, hexagon, pentagon, and rectangle. The child was handed one shape at a time and was instructed where on the sheet to glue their shape (e.g., "Let's place this one in the middle" or "This one should go below the circle" or "Let's place this one on the left"). After they glued the shape onto the page, the child was asked to name the shape. This activity was about 7 min in duration.

I Spy. On two separate days of play, children in the nonconstructive-play condition sat with an experimenter and were asked to locate specific geometric shapes on a particular page of the *I Spy Little Numbers* board book by Jean Marzollo. For example, the experimenter might say, "I spy, with my little eye, rectangles on the left-hand page. Can you find a rectangle?" The child was asked to locate three items on each page. This activity was about 6 min in duration.

Shape scavenger hunt. Children in the nonconstructive-play condition were asked to find nine, large, laminated cards with a single, colorful geometric shape. The laminated cards were placed in various locations in a room in the child's preschool. Once they located a laminated card, children were asked to name the shape and its location in the room ("What shape is that? That's right, it is a square" "Where did you find? That's right. It was in the corner of the window"). The geometric shapes on the cards were a yellow star, an aqua triangle, an orange circle, a purple rectangle, a blue octagon, a red hexagon, a green trapezoid, a blue heart, a green circle, and a blue triangle. This activity was about 7 min.

Shape memory game. On two distinct days of play activities, children in the nonconstructive-play condition played a matching game with a deck of cards. Participants were shown laminated cards with four matching pairs of distinct shapes: pink diamonds, red squares, orange pentagons, and blue octagons. The experimenter asked the child to count each side of a shape and to name it. Next, the experimenter placed the cards face down in a deck, shuffled them, and arranged the cards in two rows. Children were invited to turn over one card, name the shape, and guess which overturned card might be the shape match. Once the match was found, the experimenter would ask the child to name the shape again. This activity continued until all shape pairs had been match. This activity was 6 min in duration.

Magnetic shapes. All magnetic geometric shapes were placed on one white board while the other white board was left clear for the activity. In the nonconstructive-play condition, the experimenter asked children to find specific geometric shapes and place the shape in a specific location on the other white board of the Imagination Magnets set. For example, the experimenter would ask the child to find the yellow square among the shapes. If the child selected a different shape, the experimenter would name that shape and then ask for the first shape again. The experimenter would point out the shape if needed and remind the child of the number of sides of the shape. The child was then asked to place the shape in a specific location on the white board (e.g., "in the middle"). Children were asked to find 10 geometric shapes (e.g., two squares, one triangle, one rectangle, one large circle, two small circles, one semicircle, two arcs) and to place each in distinct locations (e.g., in the middle, under, on top, between, in the right-hand top corner, on the left side, at the bottom). This activity was about five minutes in duration.

Shape and color sort. In this play activity with the Lauri Shape and Color Sorter toy, children in the nonconstructive-play condition were shown a foam rectangular board with a horizontal row of small holes. The experimenter then placed five pegs, each a distinct color, in front of the child and asked the child to insert each peg into a distinct hole on the foam board. For example, the experimenter would instruct the child to place the green peg in the middle. Then the experimenter would ask the child to place the orange peg in the left hole and then to place the blue peg in between the green and orange peg. Once all five pegs had been inserted into the foam rectangular base, the child was given foam shapes with holes in their center. Children were asked to name each of the five shapes, a triangle, heart, square, circle, and star. Then, children were directed to place a foam shape on a specific peg (e.g., "Can you put the heart on the peg that is all the way to the left."). This activity was about 5 min in duration.

Noninterlocking puzzle. Children were guided in putting together laminated cards in a noninterlocking puzzle to form a picture with seven geometric shapes. Children were first shown a laminated picture of the completed puzzle with the seven geometric shapes, a green triangle, a red square, a light blue hexagon, an orange pentagon, a dark blue octagon, a purple rectangle, and a yellow circle. They were asked to name each shape on the sheet. Once the child had named or been told the name of each shape, they were shown 16 small, laminated squares, each depicting a portion of two shapes (e.g., part of the red square and yellow circle). The experimenter then asked the child to match their laminated square to the image on the sheet, providing scaffolding when needed. This activity was about 8 min in duration.

Across conditions, there was some degree of similarity in the play activities. For example, on the first day of play activities, children in the constructive-play condition created the origami pig and one of the laminated cards on the Imagination Magnet board set while those in the nonconstructive-play condition completed the connect-the-dots pig face and arranged shapes on the magnet board (without creating a specific pattern). On another day of play activities, children in the constructive-play condition completed a jigsaw puzzle, whereas children in the nonconstructive-play condition completed a noninterlocking puzzle in which they aligned laminated pieces to form a shape. In both conditions, the experimenter provided the spatial language most relevant to the particular activity, which often included naming geometric shapes, locations, orientations, spatial arrangements between objects, and spatial dimensions. These new activities were comparable in duration to the play activities in Experiment 1, totaling to about one hour of play across the five play sessions.

We transcribed the videos from the play activities in each condition and calculated the average number of tokens and types provided to children, distinguishing between those that were spatial and those that were nonspatial, using the same spatial language coding scheme as in Experiment 1 (Cannon et al., 2007). We compared the proportion of experimenter spatial tokens provided to children in each condition in an analysis of covariance (ANCOVA), including the average number of nonspatial experimenter tokens and the average time in the play activities as covariates. The analysis did not yield any significant effects. Children in the constructive-play condition heard an equivalent proportion of spatial language tokens (M = .179, SD = .015) as their peers in the nonconstructive-play condition (M = .182, SD = .018), $F(1, 20) = 1.84$, $p = .19$.

A comparable analysis with the proportion of experimenter spatial types in an analysis of covariance (ANCOVA) yielded only a significant effect of the average number of nonspatial experimenter types, $F(1, 30) = 2.65$, $p = .004$, $\eta_p^2 = .24$. Experimenters provided an equivalent proportion of unique spatial language (i.e., types) in each condition (M = .23, SD = .01 in the constructive-play condition and M = .21, SD = .02 in the nonconstructive play condition).

How did this input compare to the spatial language provided to children in each condition of Experiment 1? Because there were no condition effects of experimenter spatial language for Experiment 2, we combined both conditions into a single group (i.e., Experiment 2 spatial-language) to compare to each condition of Experiment 1. A 3 (Condition: Experiment 1 play-only vs. Experiment 1 spatial-language vs. Experiment 2 spatial-language) analysis of covariance (ANCOVA) on the proportion of experimenter spatial tokens with the average number of experimenter nonspatial tokens included as a covariate, yielded a significant effect of condition, $F(2, 80) = 51.94$, $p < .001$, $\eta_p^2 = .57$. Simple contrasts indicated that children in Experiment 2 heard a higher proportion of spatial tokens (M = .18, SD = .02) than those in the play-only condition of Experiment 1 (M = .12, SD = .01), $p < .001$, but not those in the spatial-language condition of Experiment 1 (M = .17, SD = .02), $p = .058$. Similarly, a 3 (Condition: Experiment 1 play-only vs. Experiment 1 spatial-language vs. Experiment 2 spatial-language) on the proportion of experimenter spatial types, with the average number of experimenter nonspatial types included as a covariate, yielded a significant effect of condition, $F(2, 80) = 61.32$, $p < .001$, $\eta_p^2 = .61$. Simple contrasts indicated that children in Experiment 2 heard a higher proportion of spatial types (M = .2196, SD = .016) than those in the play-only condition of Experiment 1 (M = .16, SD = .02), $p < .001$, but not those in the spatial-language condition of Experiment 1 (M = .22, SD = .02), $p > .70$. These results document that children in Experiment 2 heard a greater proportion of spatial tokens and types than the play-only control condition of Experiment 1, but a comparable amount and diversity of spatial language as children in the spatial-language condition of Experiment 1.

Posttest measures. Posttest measures were given one to 10 days after the last training session. Children in the nonconstructive-play condition did not differ in the number of days between the last play session and the first day of posttest measures (M = 3.71 days, SD = 2.26 days) from those in the constructive-play condition (M = 4.88 days, SD = 2.42 days), $t = 1.47$, $p = .15$. Children were given the alternate forms of the CMTT and PRT at posttest but the same version of the Spatial Vocabulary Assessment. The PPVT-IV was given on the final day of posttest testing, a contrast to Experiment 1 in which this measure was administered at baseline. Finally, children completed two additional measures on a separate day of assessments. These measures are not included in the present study because they are the focus of a distinct research question.

Results

Descriptive statistics for the PPVT and the baseline and posttest assessments of the PRT and Spatial Vocabulary Assessment are listed in Table 5. Preliminary analyses did not yield any significant effects of child sex on the baseline or posttest measures of Spatial Vocabulary, $ps > .24$, or general vocabulary (i.e., the PPVT- IV),

Table 5
For Each Condition of Experiment 2, the Mean (and Standard Deviation) of Children's Percent Correct on Each Measure

	Constructive-play condition		Nonconstructive-play condition	
Condition	Baseline	Posttest	Baseline	Posttest
PRT				
M (*SD*)	.431 (.187)**	.495 (.190)**	.377 (.084)**	.422 (.146)**
95% CI	[.335, .527]	[.398, .593]	[.334, .421]	[.347, .497]
Spatial vocabulary				
M (*SD*)	.609 (.164)	.676 (.149)	.598 (.143)	.665 (.142)
95% CI	[.524, .693]	[.599, .753]	[.522, .674]	[.592, .738]
PPVT (single administration)				
M (*SD*)	59.6 (28.4)		54.0 (28.2)	
95% CI	[45.0, 74.1]		[39.5, 68.5]	

Note. PPVT = Peabody Picture Vocabulary Test.
** $p < .01$.

$p > .63$; however, there was a significant effect of child sex on the baseline measure of the PRT, $t(31) = 2.38$, $p = .02$. Child sex was collapsed in subsequent analyses for those measures for which there was no effect of child sex but included in the analyses for the PRT.

Relation between measures. We explored the relation between our measures and child age (see Table 6). Age was significantly related to children's baseline and posttest performance on their posttest spatial vocabulary. In addition, children's general vocabulary (PPVT) was significantly related to their baseline and posttest performance on the PRT as well as their spatial vocabulary at posttest. To control for both age and general vocabulary, we conducted partial correlations that controlled for these two variables. Children's baseline and posttest performance on the PRT were significantly related. In addition, children's posttest PRT was significantly related to their posttest spatial vocabulary.

General vocabulary. An analysis of variance (ANOVA) did not yield an effect of condition on children's percentile on the PPVT-IV, $F < 1$, $p = .93$.

Spatial vocabulary. An analysis of covariance (ANCOVA) examined the effect of condition on children's posttest spatial vocabulary, with their baseline spatial vocabulary, the number of experimenter nonspatial tokens, and the average duration of the play activities as covariates. The analysis did not yield any significant effects. There was no effect of condition, $F < 1$, $p = .93$. As a group, children demonstrated significant gains on our measure of spatial vocabulary from baseline to posttest, paired sample $t(32) = 2.05$, $p = .048$. With the exception of two participants, all children ($n = 32$) in the sample demonstrated gains in their spatial vocabulary, a binomial distribution significantly greater than chance, $p < .001$.

Mental rotation. One girl in the nonconstructive play condition scored three standard deviations below the mean of the sample and was omitted from the present analysis. The effect of condition and child sex on children's percent correct at posttest on the PRT was examined in an ANCOVA, with their percent correct at baseline, the average number of experimenter's nonspatial tokens, and average play time included as covariates. The analysis yielded only a significant effect of baseline performance, $F(1, 26) = 28.67$, $p < .001$, $\eta_p^2 = .52$. As a group, children demonstrated significant gains from baseline to posttest on the PRT, paired $t(32) = 3.88$, $p < .001$. This result held for children in the nonconstructive-play condition and constructive-play condition separately, paired $t(15) = 2.93$, $p = .01$ and paired $t(16) = 2.52$, $p = .02$, respectively (see Figure 4).

Comparison with the control condition of Experiment 1. Did children in Experiment 2 demonstrate significant gains in their mental rotation relative to children in the play-only condition of Experiment 1, who did not hear spatial labels? Children in Experiment 2 scored significantly lower at baseline on the mental rotation task than those in the play-only condition of Experiment 1, $t(56) = 2.87$, $p < .01$. They also scored lower on the measure of spatial vocabulary at baseline, $t(55) = 2.95$, $p < .01$, and the PPVT general vocabulary measure, $t(54) = 3.07$, $p < .01$. Given that children in Experiment 2 had significantly lower baseline measures, we computed a percent change on the PRT rather than compare performance on posttest PRT scores. The percent change was calculated by subtracting baseline from posttest scores on the PRT to generate a difference score. Next, this difference score was divided by the baseline score and then multiplied by 100, yielding our percent change on the mental rotation task. A 2 (Experiment: 1 play-only control vs. 2 spatial

Table 6
Correlations Between Child Age, Baseline, and Posttest Measures for the PRT and Spatial Vocabulary, and General Vocabulary in Experiment 2

Age, task	1	2	3	4	5	6
1. Child age		.168	.276	.305	.222	.388*
2. BL PRT			.226	.461*	.803*	.498*
3. BL spatial vocabulary		.144		.216	.047	.145
4. General vocabulary					.431*	.612*
5. PT PRT		.760*	−.079			.534*
6. PT spatial vocabulary		.311	−.050		.363*	

Note. The correlations between child age, baseline, and posttest measures for the PRT and spatial vocabulary, and general vocabulary are listed in the top half of Table 6. Partial correlations, controlling for child age and general vocabulary (PPVT-IV), are listed below the diagonal. BL PRT = baseline PRT; BL = baseline; PPVT-IV = Peabody Picture Vocabulary Test IV; PT PRT = posttest PRT; PT = posttest.
* $p < .05$.

language) analysis of variance (ANOVA) on the PRT percent change yielded a significant effect of Experiment, $F(1, 54) = 5.78$, $p = .02$, $\eta_p^2 = .10$. Children in Experiment 2 ($M = 21.19$, $SD = 33.88$) demonstrated a significantly higher percent change than children in the play-only condition of Experiment 1 ($M = -5.68$, $SD = 30.98$). This result shows that children in Experiment 2 improved more in mental rotation from baseline to posttest than those in the play-only control condition of Experiment 1.

Does the proportion of spatial language account for percent change in mental rotation across both experiments? We conducted a series of regression analyses to assess whether experiment and the proportion of experimenter spatial tokens predicted percent change on children's mental rotation when results from all children in both experiments were examined. Children in Experiment 2 also scored significantly lower on the general vocabulary measure than children in the spatial-language condition of Experiment 1, $t(58) = 4.14$, $p < .001$. For this reason, child general vocabulary was included as a variable in the regression model. A model with baseline PRT performance, child age, child general vocabulary, experiment (1 vs. 2), and the average number of experimenter nonspatial tokens was significant, $F(5, 75) = 3.47$, $p = .007$, with an adjusted R^2 of .13. Only children's baseline score, $\beta = -.42$, $t = -3.83$, $p < .001$, significantly predicted children's percent change on the PRT in this first model. A second model that added proportion of experimenter spatial tokens accounted for a significant amount of the variance, $\Delta F(1, 74) = 12.36$, $p = .001$, with a ΔR^2 of .12 and an adjusted R^2 of .25. In this model, children's baseline PRT, $\beta = -.34$, $t = -3.27$, $p = .002$, the proportion of experimenter spatial tokens, $\beta = .43$, $t = 3.52$, $p = .001$, and experiment, $\beta = -.36$, $t = -2.16$, $p = .03$, accounted for a significant amount of the variance in children's percent change in their PRT performance.

Analyses with the proportion of experimenter spatial types yielded similar results. Model 1 with children's baseline PRT, age, child general vocabulary, experiment, and the average number of experimenter nonspatial types was significant, $F(5, 75) = 3.73$, $p = .005$, with an adjusted R^2 of .15. Only children's baseline PRT, $\beta = -.42$, $t = -3.91$, $p < .001$, predicted percent change in the PRT scores in this first model. Model 2, which added the proportion of experimenter spatial types, yielded a significant model, $\Delta F(1, 74) = 9.30$, $p = .003$, with a R^2 change of .09 and an adjusted R^2 of .23. Children's baseline PRT, $\beta = -.35$, $t = -3.31$, $p = .001$, the proportion of experimenter spatial types, $\beta = .34$, $t = 3.05$, $p = .003$, and experiment, $\beta = -.34$, $t = -2.11$, $p = .04$, predicted a significant amount of the variance in the percent change scores on the PRT.

In each set of regression models, children's baseline performance and the proportion of spatial language tokens or types accounted for a significant amount of the variance in children's percent change on the mental rotation task. In addition, the sample of children (Experiment 1 vs. 2) also explained the variance on children's percent change but only when experimenter spatial language was included in the model. Without the proportion of experimenter spatial language in the model, the experiment in which children participated was unrelated to their change in mental rotation performance. With the proportion of experimenter spatial language in the model, experiment contributes to the variance in the percent change in mental rotation because children in Experiment 2 had more modest gains in the percent change in mental rotation scores relative to their peers in the spatial-language condition of Experiment 1.

Discussion

Children in both conditions of Experiment 2 demonstrated significant improvements in their spatial vocabulary and mental rotation. These children also demonstrated greater gains in their mental rotation relative to the play-only control group from Experiment 1. After five sessions of exposure to spatial language during play rich with spatial features, children significantly improved on the mental rotation task relative to baseline. There was no evidence that the constructive nature of the spatial play interacted with spatial language input in promoting children's mental rotation. Rather, the results suggest that the more critical element for boosting this spatial skill was enriching children's exposure to spatial language more so than pairing this input with a particular type of constructive play.

General Discussion

The current experiments provide empirical support for a causal link between children's experience with spatial language during spatial play and advances in their mental rotation. After five weekly or semiweekly play sessions, children who heard more spatial language demonstrated significant gains in mental rotation relative to peers who heard less spatial language. The total amount of spatial language input in the present studies was modest. Each 12-min play session summed to about one hour across all play sessions. Nonetheless, these relatively brief bursts of spatial language, embedded in adult-guided spatial activities, yielded gains in mental rotation that were maintained for at least one or more days. The present results provide the first experimental evidence that children's experience with spatial language during spatial play can benefit the nontrained spatial skill of mental rotation, complementing naturalistic studies of parent and child spatial language use and preschool children's spatial skills (Pruden et al., 2011).

The present results outline several new insights into the role of spatial language in the development of spatial skills. First, exposure to spatial language benefits the growth of mental rotation, even though this spatial skill was seemingly unrelated to the spatial language provided. The present studies also highlight how spatial labels need not be tailored to or be temporally contiguous with a mental rotation task to promote children's performance on this task. Rather, embedding spatial language into adult-child interactions can generalize to enhance this aspect of children's spatial thinking. Second, experimenter spatial language uniquely contributed to advances in mental rotation across both constructive and nonconstructive activities, suggesting that infusing spatial language when spatial information is present may be as effective as doing so when there is a constructive element paired with the spatial information, as in block building and puzzle play (Ferrara et al., 2011; Levine et al., 2012). That is, spatial language does not depend on its pairing with constructive play to promote mental rotation. Third, our results replicated across a middle- and low-income sample of 4-year-old children, documenting that the present results are not restricted to children with a particular proficiency in spatial or language skills. Children enrolled in Head Start had lower mental rotation scores and scored lower on the spatial vocabulary and general vocabulary measures than their middle-income peers but nonetheless, demonstrated significant gains from baseline to posttest. Finally, children's experience with spatial language over multiple play sessions yielded durable gains in mental rotation, extending the time frame by several days over which previous studies have documented maintenance of the facilitative effects of spatial language (e.g., Dessalegn & Landau, 2013; Loewenstein & Gentner, 2005). In a few cases, some of our participants were not assessed on their spatial skills until almost two weeks after the final play session. In sum, the present results advance our current conceptions of how spatial language can promote aspects of spatial thinking and the durability of this facilitation. The results also suggest that increasing the amount and quality of spatial language input to 4-year-old children may be a viable approach for enhancing the development of mental rotation.

Children's exposure to spatial language accounted for significant gains in mental rotation across two samples of children, the middle-income sample in the spatial-language condition of Experiment 1 and the Head Start sample in Experiment 2. Even so, the results of the regression analyses indicated that the gains for our sample of children in Head Start in Experiment 2 were more modest compared with the middle-income sample of children in Experiment 1. There are a number of factors that may have contributed to this result. Children in Head Start scored lower at baseline on the spatial and language measures than children in Experiment 1. It may be that children with more limited vocabularies, or those who struggled more with the spatial task at baseline, may not benefit to the same degree as children with stronger language or spatial skills at the start of the study. Alternatively, the more modest increase in mental rotation for the children in Head Start may be attributable to any combination of factors that differed across the two experiments, including the specific spatial play activities and the spacing of the training sessions. Without including middle- and low-income children in the same sample, it is not possible to rule out the contribution of the design and activities differences across the two experiments nor to draw any conclusions for the stronger effect of our spatial language manipulation in the middle- than low-income preschool children.

As children's exposure to spatial language tokens and types increased, so did their posttest spatial performance. Both the amount and diversity of experimenter spatial language uniquely contributed to the advances in mental rotation. This result emerged whether the results of Experiment 1 were considered alone or in combination with the results of Experiment 2. Results from exploratory analyses in Experiment 1 suggest that spatial language was especially effective for enhancing children's performance on the more difficult items, those rotated at higher angles from an upright orientation. Consistent with other experimental studies (e.g., Loewenstein & Gentner, 2005), the effects of the spatial language on children's performance were most evident when the task was challenging. In contrast, for items at lower angles of rotation, children's use of spatial language, and not the experimenter spatial language, accounted for a significant amount of the variance in performance, consistent with the many studies linking children's spatial vocabulary to their spatial performance (Ankowski et al., 2012; Balcomb et al., 2011; Gentner et al., 2013; Hermer-Vazquez et al., 1999; Miller et al., 2017; Park & Casasola, 2017; Piccardi et al., 2015; Pruden et al., 2011; Sims & Gentner, 2008). This pattern of results is intriguing and suggests that child spatial language during the play activities related to mental rotation performance in distinct and perhaps complementary ways than the experimenter spatial language. Yet, this claim remains to be further explored, particularly with a measure of mental rotation that includes sufficient number of items at higher versus lower angles of rotation.

How did children's experience with spatial language translate to improvements in mental rotation, particularly given that the spatial language did not readily align with this task? Gentner (2003, 2016) has proposed that spatial language can enhance spatial thinking through the process of comparison. Providing spatial labels for geometric shapes and the spatial features of objects may have significantly strengthened children's representations for this spatial information. As a result, children may have more easily shifted attention away from nonspatial features, such as color, to the spatial information of object shape. This attentional shift to object shape, particularly over nonspatial features, may have motivated the use of a more holistic strategy, which has been linked to stronger performance on the mental rotation task in both adults and children (e.g., Boone & Hegarty, 2017; Tzuriel & Egozi, 2010). Akin to how weekly training of object labels based on shared shape facilitated infants' acquisition of object names (Smith et al., 2002), children's cumulative experience with shape names may have enhanced their representations for geometric shape and by doing so, bolstered their ability to mentally rotate objects.

The results document that the amount of experimenter spatial language accounted for the same amount of variance as the experimenter's lexical diversity of spatial language on children's mental rotation, indicating that both spatial language quantity and quality contributed to advances in children's mental rotation. Just as language quality predicts the growth of children's language skills (e.g., Rowe, 2012), spatial language quality in the present study predicted gains in mental rotation performance. Also worth considering is how the context of the spatial language may have contributed to the gains in mental rotation. Providing such a diverse array of spatial labels (e.g., geometric shapes, object placement, dimensions, spatial features, spatial relations, orientation, and locations) infused within an equally diverse array of play activities may have promoted children's representations of spatial information at a broader level, beyond simply bolstering the spatial information that was labeled. Increasing the diversity and amount of spatial labels, and using these labels across diverse examples of these referents in the play activities, may have promoted a greater level of abstraction, analogous to how increasing the number and variability of exemplars in a categorization task promotes broader and more abstract generalization (e.g., Casasola & Park, 2013; Oakes, Coppage, & Dingel, 1997; Perry, Samuelson, Malloy, & Schiffer, 2010). In other words, it is not only the exposure to the spatial language but also the diverse contexts in which this input is provided that may have promoted children's spatial thinking, a possibility also in line with Gentner's (2003, 2016) assertions of the central role of comparison in bolstering spatial representations. With stronger spatial representations, children may have been able to dedicate more cognitive resources to spatial processing, explaining why exposure to such a diverse array of spatial language generalized to promote mental rotation.

Other candidate mechanisms for the obtained results may have to do with the alignment of skills across domains or the time schedule of the trainings. Providing spatial language as children manipulated geometric shapes and Lego bricks or folded paper made it possible for children to align their actions and attention to the labeled spatial information. The co-occurrence among these inputs may have created a synergy that was perfectly pitched to bolster the effect of the spatial labels on children's spatial thinking. Similarly, distributing the play trainings over several weeks may also have contributed to the present results. There is a rich literature documenting the benefits of distributed over massed learning for memory and generalization (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Childers & Tomasello, 2002; Vlach, 2014; Vlach, Sandhofer, & Kornell, 2008). Vlach and Sandhofer (2012) noted that children who received instruction once a week demonstrated stronger generalization than those who had massed learning in a single day. Spacing the trainings in the present experiments over three to four weeks may have contributed to the transfer of spatial language to promote performance on the nontrained task. In sum, there may be a number of pathways that explain how children's experience with spatial language during spatial play generalized to promote mental rotation, a spatial skill that was nontrained. As more of these pathways converge, children's spatial thinking is more likely to benefit from spatial language exposure.

Unfortunately, our attempt to create an abbreviated form of a mental transformation task compromised the reliability of the full scale. As such, we are unable to draw conclusions of whether our manipulation in spatial language input would extend to other types of spatial skills or if the present results are specific to mental rotation. Given the transfer effects reported in other training studies of spatial skills (e.g., Uttal et al., 2013), we posit that increasing children's exposure to spatial language may generalize beyond mental rotation to other types of spatial skills. However, this claim remains to be empirically tested. At a theoretical level, exploring the range of spatial skills that are shaped by spatial language is pivotal for defining the boundaries of how spatial language can promote spatial thinking.

There are a number of other issues that remain to be addressed in future work. An important extension of the present work is to explore its replicability across distinct settings. Our focus was on establishing a causal link between spatial language input and children's spatial skills. With this goal in mind, we used experimenters who would adhere to our manipulation in spatial language input. In considering the translational value of the present findings, a critical next step is assessing replicability with caregivers or preschool teachers across a broader range of social and play activities. It is not clear if the same benefits would emerge with play activities that were less spatial or in a group setting rather than in one-on-one interactions with an adult. It will also be insightful to document whether the play experiences during the training shaped children's play preferences outside of our study. We did not ask parents to report on their child's play as they participated in the study, but an interesting question is whether our study influenced play preferences.

Although many studies document sex differences in spatial skills (e.g., Lauer & Lourenco, 2016; Levine et al., 1999), there was no effect of child sex in our manipulation of spatial language (Experiment 1) or in our manipulation of the constructive nature of the spatial play activities (Experiment 2). This result is perhaps not surprising given that both sexes in our studies engaged in the same play activities and were provided with the same amounts of spatial language. In previous studies, sex differences in children's spatial language has been traced to parent spatial language, with parents of male children providing more spatial language than parents of female children (Pruden & Levine, 2017). Our results suggest that when male and female preschool children engage in the same spatial activities and are provided with equivalent amounts of spatial language, there are no sex differences in their improvement in spatial skills. It may be that because the activities were equally appealing to both the male and female children, children engaged in the play activities with the same enthusiasm and consequently, benefited equally from the spatial language input. Our manipulation may not have been as effective for both sexes if the activities had been more appealing to one gender (Coyle & Liben, 2018). Of course, sex differences in spatial performance do not always emerge (Constantinescu, Moore, Johnson, & Hines, 2018) and the present results may simply reflect that both male and female 4-year-old children's mental rotation can benefit when their spatial play is accompanied by generous amounts and diversity of spatial language.

The present results represent a first step in understanding how spatial language can exert an influence on spatial cognition. They highlight the causal effect of spatial language input on one type of children's spatial skills, mental rotation. Exposure to a broad array of spatial language, embedded during brief interactions in diverse spatial play activities, promoted children's spatial thinking, yielding gains in mental rotation that remained evident over several days. The results break new ground by documenting that spatial language can transfer to promote the nontrained spatial skill of mental rotation. One aspect of the present results that merits further exploration is whether similar results would have emerged with a more focused array of spatial terms or a more restricted set of spatial activities (e.g., a focus on Lego building). Our intuition is that it was the diversity of labels paired with an equally diverse array of activities that made it possible for spatial language to benefit children's mental rotation. Restricting the diversity in either the spatial language or their contexts may not be as effective as the approach in the present studies. We also posit that providing this exposure in relatively brief intervals distributed over several weeks may have contributed to the transfer to the nontrained spatial task. Yet, these claims remain to be systematically tested. Additional studies that vary the diversity of spatial language, the context in which that language is provided, and timing of the exposure will be critical in not only replicating the present results but also for elucidating the mechanisms by which spatial language promotes children's mental rotation.

Context Paragraph

Pruden et al. (2011) documented a relation between children's spatial language use as toddlers and their spatial skills as preschoolers. Their results raised the exciting possibility that children who hear more spatial language develop stronger spatial skills, a relation mediated by children's spatial language. However, their study left open the possibility that other variables may have contributed to the relation between children's spatial language and their spatial skills. Inspired by their findings, we directly tested the effect of increasing preschool children's experience with spatial language on the development of their spatial skills. We also sought to isolate the effect of spatial language input from the effect of constructive play. For this reason, we held constructive play constant while manipulating spatial language input in the first experiment, and we manipulated constructive play while holding spatial language input constant in the second experiment. These studies are the first to explore the impact of spatial language across spatial play contexts and over multiple sessions. The present results uniquely document that increased exposure to spatial language during spatial play generalizes to benefit mental rotation, yielding new insights into the role of spatial language in the development of one aspect of children's spatial cognition.

References

Ankowski, A. A., Thom, E. E., Sandhofer, C. M., & Blaisdell, A. P. (2012). Spatial language and children's spatial landmark use. *Child Development Research, 2012,* 1–14. http://dx.doi.org/10.1155/2012/427364

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20,* 327–343. http://dx.doi.org/10.1007/bf00287729

Balcomb, F., Newcombe, N. S., & Ferrara, K. (2011). Finding where and saying where: Developmental relationships between place learning and language in the first year. *Journal of Cognition and Development, 12,* 315–331. http://dx.doi.org/10.1080/15248372.2010.544692

Berneiser, J., Jahn, G., Grothe, M., & Lotze, M. (2018). From visual to motor strategies: Training in mental rotation of hands. *NeuroImage, 167,* 247–255. http://dx.doi.org/10.1016/j.neuroimage.2016.06.014

Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1005–1019. http://dx.doi.org/10.1037/xlm0000370

Cannon, J., Levine, S., & Huttenlocher, J. (2007). A system for analyzing children and caregivers' language about space in structured and unstructured contexts. *Spatial Intelligence and Learning Center (SILC) technical report*. Retrieved from https://www.silc.northwestern.edu/?s=Spatial+Intelligence+and+Learning+Center+%28SILC%29+technical+report

Carey, S. (2009). *The origins of concepts*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195367638.001.0001

Casasola, M. (2005). Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. *Developmental Psychology, 41,* 183–192. http://dx.doi.org/10.1037/0012-1649.41.1.183

Casasola, M., & Park, Y. (2013). Developmental changes in infant spatial categorization: When more is best and when less is enough. *Child Development, 84,* 1004–1019. http://dx.doi.org/10.1111/cdev.12010

Casey, B. M., Erkut, S., Ceder, I., & Young, J. M. (2008). Use of storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology, 29,* 29–48. doi:10.1016/j.appdev.2007.10.005

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380. http://dx.doi.org/10.1037/0033-2909.132.3.354

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 38,* 967–978. http://dx.doi.org/10.1037/0012-1649.38.6.967

Constantinescu, M., Moore, D. S., Johnson, S. P., & Hines, M. (2018). Early contributions to infants' mental rotation abilities. *Developmental Science, 21*. http://dx.doi.org/10.1111/desc.12613

Coyle, E. F., & Liben, L. S. (2018). Gendered packaging of a STEM toy influences children's play, mechanical learning, and mothers' play guidance. *Child Development*. Advance online publication. http://dx.doi.org/10.1111/cdev.13139

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

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

Dunn, M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test – 4*. Circle Pines, MN: AGS.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

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. http://dx.doi.org/10.1111/j.1751-228X.2011.01122.x

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–236). Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/4117.003.0015

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist, 71,* 650–657. http://dx.doi.org/10.1037/amp0000082

Gentner, D., Ozyürek, A., Gürcanli, O., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition, 127,* 318–330. http://dx.doi.org/10.1016/j.cognition.2013.01.003

Hart, B., & Risley, T. R. (1995). *Meaningful difference in the everyday experiences of young American children*. Baltimore, MD: Paul H. Brooks Publishing.

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education, 4,* 60–68. http://dx.doi.org/10.1016/j.tine.2015.05.001

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

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

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. sample. *Psychological Science, 26,* 302–310. http://dx.doi.org/10.1177/0956797614563338

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

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35,* 940–949. http://dx.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. http://dx.doi.org/10.1037/a0025913

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science, 16,* 841–845. http://dx.doi.org/10.1111/j.1467-9280.2005.01623.x

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

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (Vol. 2). Mahwah, NJ: Erlbaum.

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

Miller, H. E., Vlach, H. A., & Simmering, V. R. (2017). Producing spatial words is not enough: Understanding the relation between language and spatial cognition. *Child Development, 88,* 1966–1982.

Mix, K. S., Levine, S. C., Cheng, Y. L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General, 145,* 1206–1227. http://dx.doi.org/10.1037/xge0000182

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

Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator, 34,* 29–35.

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

Oakes, L. M., Coppage, D. J., & Dingel, A. (1997). By land or by sea: The role of perceptual similarity in infants' categorization of animals. *Developmental Psychology, 33,* 396–407. http://dx.doi.org/10.1037/0012-1649.33.3.396

Öostermeijer, M., Boonen, A. J. H., & Jolles, J. (2014). The relation between children's constructive play activities, spatial ability, and mathematical word problem-solving performance: A mediation analysis in sixth-grade students. *Frontiers in Psychology, 5,* 782. http://dx.doi.org/10.3389/fpsyg.2014.00782

Oudgenoeg-Paz, O., Leseman, P. P., & Volman, M. C. (2015). Exploration as a mediator of the relation between the attainment of motor milestones and the development of spatial cognition and spatial language. *Developmental Psychology, 51,* 1241–1253. http://dx.doi.org/10.1037/a0039572

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108,* 155–184. http://dx.doi.org/10.1016/j.cognition.2008.02.007

Park, Y., & Casasola, M. (2017). The impact of object type on the spatial analogies in Korean preschoolers. *Cognitive Psychology, 94,* 53–66. http://dx.doi.org/10.1016/j.cogpsych.2017.02.001

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally. Exemplar variability supports higher-order generalization and word learning. *Psychological Science, 21,* 1894–1902. http://dx.doi.org/10.1177/0956797610389189

Piccardi, L., Palermo, L., Bocchi, A., Guariglia, C., & D'Amico, S. (2015). Does spatial locative comprehension predict landmark-based navigation? *PLoS ONE, 10,* e0115432. http://dx.doi.org/10.1371/journal.pone.0115432

Pruden, S. M., & Levine, S. C. (2017). Parents' spatial language mediates a sex difference in preschoolers' spatial-language use. *Psychological Science, 28,* 1583–1596. http://dx.doi.org/10.1177/0956797617711968

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. http://dx.doi.org/10.1111/j.1467-7687.2011.01088.x

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 12116–12120. http://dx.doi.org/10.1073/pnas.0914044107

Quaiser-Pohl, C. (2003). "The mental cutting test Schnitte" and the Picture Rotation Test – two new measures to assess spatial ability. *International Journal of Testing, 3,* 219–231. http://dx.doi.org/10.1207/S15327574IJT0303_2

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science, 19,* 1067–1070. http://dx.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. http://dx.doi.org/10.1111/infa.12033

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83,* 1762–1774. http://dx.doi.org/10.1111/j.1467-8624.2012.01805.x

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703. http://dx.doi.org/10.1126/science.171.3972.701

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

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—But not circular ones—Improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101,* 545–560. http://dx.doi.org/10.1037/a0014239

Sims, N., & Gentner, D. (2008). Spatial language and landmark use: Can 3-, 4-, and 5-year-olds find the middle? *Proceedings of the Cognitive Science Society*. Retrieved from https://escholarship.org./uc/item/5fk7p56q

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13,* 13–19. http://dx.doi.org/10.1111/1467-9280.00403

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22,* 996–1013. http://dx.doi.org/10.1002/acp.1420

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development, 81,* 1417–1430. http://dx.doi.org/10.1111/j.1467-8624.2010.01482.x

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

Vander Heyden, K. M., Huizinga, M., & Jolles, J. (2017). Effects of a classroom intervention with spatial play materials on children's object and viewer transformation abilities. *Developmental Psychology, 53,* 290–305. http://dx.doi.org/10.1037/dev0000224

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology, 126,* 37–51. http://dx.doi.org/10.1016/j.jecp.2014.02.012

Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives, 8,* 163–168. http://dx.doi.org/10.1111/cdep.12079

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development, 83,* 1137–1144. http://dx.doi.org/10.1111/j.1467-8624.2012.01781.x

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109,* 163–167. http://dx.doi.org/10.1016/j.cognition.2008.07.013

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

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology, 30,* 679–689. http://dx.doi.org/10.1037/0012-1649.30.5.679

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

Xu, F. (2016). Preliminary thoughts on a rational constructivist approach to cognitive development: Primitives, symbols, learning, and thinking. In D. Barner & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 11–28). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190467630.003.0002

Appendix

Sample of Experimenter Language Across Activities and Conditions in Each Experiment

Experiment 1		
Activity	Spatial-language condition	Play-only condition
Origami Pig	"We are going to bring this left corner to this right corner and make a vertical line."	"Can you take this point and fold it to this point? You gotta fold it all the way over there"
Origami Whale	"Now we are going to take the right-hand side and fold it to that vertical line." "Fold it all the way up to make a triangle."	"So we fold it like this. Can you fold it down? And now you open it up." "The tricky part here is that this has to line up with that."
Magnetic Board and Shapes	"Put the blue one on the right and yellow one on the left." "You want to use a big triangle?" "What goes under it?"	"Where shall we place it? Maybe move it a little more to here." "You want to use that piece there?" "Which one goes next?"
Shape Matching	"See these pieces. Which one of these four shapes do you think it makes? The triangle, the pentagon, the hexagon, or the octagon"? [Experimenter points to each shape.]	"See these pieces. Which one of these four shapes do you think it makes? This one here, or the one here, or this one here, or that one?" [Experimenter points to each shape.]
Lego Build	"Can you find this rectangle?" "Yes, those are both rectangles, but this is a little taller." "We are going to put that on top of the red one and next to the gray one."	"Can you find this piece?" "We have all these pieces. Let's look to see which one we need." "Let's find this shape first; and now we need these two black pieces."

Experiment 2			
Constructive-play condition		Nonconstructive-play condition	
Activity	Spatial language	Activity	Spatial language
Origami Pig	"Let's make a horizontal line by folding the paper up." "Can you fold this top corner down to the bottom?" "Now we have a horizontal line, right?" "Let's just fold this one down into a triangle."	Connect the Dots	"We are going to make a line from two to three." "Do you know what shape this is?" A circle! Good job." "It has no edges. It goes all the way around." "You made a pig with all different shapes."
Magnetic Board and Shapes	"Those are called arcs, can you say arcs?" "The green's on top, on top of what?" "Let's start with this big blue square."		"Can you put this at the top of the board here? Now place this one in the middle." "What shape do you wanna do next?" "Can you put it under, under the semi-circle?"
Block Building	"Can you make the rectangle with those two pieces?" "How do we make a rectangle with two squares?"	Shape Scavenger Hunt	"What shape is this? It's a hexagon!" "Where did you find it? Yes, under the chair."

(*Appendix continues*)

Appendix (*continued*)

Experiment 2			
Constructive-play condition		Nonconstructive-play condition	
Activity	Spatial language	Activity	Spatial language
Origami Whale	"We're gonna fold this left corner all the way to that right corner." "What shape did we just make?" "We're gonna fold this corner into the middle and have it make a horizontal line."	Shape Sorting Toy	"Can you put this piece all the way to the right side?" "Can you put this green piece to the left?" "Now can you put this red piece in the middle?" "Now can you put the orange piece between the blue and the red?"
Magnatiles	"Let's see if you can put the triangles together to make a square." "There is a square in the middle." "Can you put them together to make a square that will be on top?" "So you used the triangles to make a square. That was perfect!"	Shape Collage	"You're putting it in the left corner." "I'm gonna put the heart under the triangle just like that." "Is that on top or the bottom?" "I'm gonna put the diamond in the corner."
Lego Build 1	"So you put down the square" "What do we have to put on the square?" "I think we have to find this green rectangle."	Shape Memory Game	"You got your first match! What shape is this?" "Where's the other square?" "A triangle! Where's another triangle."
Jigsaw Puzzle	"That goes on top." "Look, they both line up with the straight edge, right?" "This piece comes out and this piece goes in so they fit together."	Non-interlocking puzzle	"Put that on the bottom of that one." "Right here in the corner." "Can you put that right next to it on the left."
Lego Build 2	"Do you see these shapes over here? Can you put them on top of the light green rectangle? I think they go all the way to the edge."	I Spy book	"What do I spy next to the rectangle?" "I spy a triangle. Can you spy another triangle?" "Is it next to the square?"

Received July 21, 2017
Revision received September 1, 2019
Accepted September 12, 2019 ■