Published in final edited form as:

J Mem Lang. 2018 August ; 101: 136–152. doi:10.1016/j.jml.2018.04.002.

Language unifies relational coding: The roles of label acquisition and accessibility in making flexible relational judgments

Nicole M. Scott^{1,*} and Maria D. Sera²

¹Center for Cognitive Sciences, University of Minnesota, 75 East River Road, Minneapolis, MN 55417, USA

²Institute of Child Development, University of Minnesota, 51 East River Road, Minneapolis, MN 55455, USA; 612-624-2856

Abstract

Language is likely structuring spatial judgments, but how it achieves this is not clear. We examined the development of relative, spatial judgments across verbal and nonverbal tasks of *above*, *below*, *right* and *left* in children between the ages of 5 and 10 years. We found that the verbal ability to make *above/below* judgments preceded verbal *right/left* judgments and all nonverbal judgments. We also found that only when the labels were *accessed* – as opposed to only having been acquired – did children's nonverbal performance improve. Our findings further indicate that accessing the correct term was not needed for enhanced performance. The results suggest that accessing language unifies different instantiations of a relation into a single representation.

Keywords

spatial relations; entrenchment in language development; cognitive development; relational terms; language acquisition; language accessibility

Introduction

The ability to make relational judgments is central to human intelligence because it promotes abstract levels of reasoning and problem solving (Gentner, 2003, 2010; Halford, Wilson & Philips, 2010). For example, relational knowledge is fundamental in spatial reasoning, which in turn is critical for math and science skills (National Research Council, 2006). By some views, knowledge of language plays a critical role in relational processing.

Corresponding author: Maria D. Sera.

*Dr. Scott is currently at Grand View University, Department of Psychology, 1200 Grand View Ave, Des Moines, IA 50316

Declarations of interest: none

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Yet, despite the evidence that language affects relational processing, it remains unclear exactly *how* language might be helping. It is largely believed that as each individual term is learned, it instantly enhances performance across different instantiations of that relation *automatically* without conscious effort, as seen in some Stroop tasks (see Diamond, Kirkham, & Amso, 2002, for discussion). In this case, once a term is learned, it is automatically retrieved and applied in the appropriate contexts. Another possibility, however, is that language *use* is initially needed and improvement in nonverbal judgments is only observed when the labels are *accessed*. In this case, the appropriate terms are available but are not always activated or applied to the task at hand. Only after the terms are entrenched or well-learned would they, then, affect performance effortlessly across different tasks or instantiations of the relation.

The goal of this paper is to better understand the role of language in making dynamic relational judgments. We do this in two ways. First, we examine developing knowledge of a broader set of relations across a broader age span than has been previously studied, both verbally and nonverbally. This enables us to examine the role of language *acquisition* in the ability to make correct relational judgments. Second, we manipulated the *accessibility* of language to examine whether language and its *use* are causal factors in any improvement of relational processing. While several studies have found that language knowledge precedes enhanced performance across other relational tasks, because these studies are largely correlational, it is not clear whether language or an unstudied factor (that is also correlated with language development) is driving the improvement. If language is causing the improvement, however, and not some third factor, then “dialing up” the strength of language should improve performance in nonverbal relational judgment tasks. If another factor is solely responsible for the improvement, then dialing up the role of language should not affect performance. A final issue that we studied was whether the use of any term would suffice, or whether coding with the precise relational terms was necessary for performance to improve.. This study, thus, fills important gaps in the evidence on the role of language in the development of the ability to make relational judgments.

Past Evidence on the Role of Language in Relational Thought

In trying to understand the development of relational knowledge, many researchers have turned to language as a causal factor, but the current evidence regarding the role of language in relational knowledge is mixed. There is some cross-linguistic evidence indicating that language can structure spatial representations. For example, Hespos and colleagues (Hespos & Spelke, 2004; Hespos & Piccin, 2009) found that 5-month-old infants being raised in an English-speaking environment were sensitive to *tight* versus *loose* distinctions - a distinction not marked in English, but that is marked in other languages, like Korean. In these studies, English-reared infants behaved like Korean adults, and unlike English adults, in recognizing the *tight-loose* distinction, suggesting that language leads English-speaking adults to ignore the distinction. In other cross-linguistic work on the topic, Levinson (1996) showed that preferred frames of reference are determined by language: whether a language codes spatial information in terms of *north-south* (cardinal) or *left-right* (relational) coordinates effectively determines how speakers represent object locations (see also, Levinson 2003). Other studies, however, have shown that individuals can override their preferred, native

frame of reference for less preferred reference points, in certain situations (Li & Gleitman, 2007). However, such flexibility has not always been found (e.g., Haun, Rapold, Janzen, & Levinson, 2011).

When studying development, or what comes first, considerable evidence suggests that acquiring labels consistently precedes the ability to encode and remember relative object locations (e.g., Pyers et al., 2010; Loewenstein & Gentner, 2005). For example, Loewenstein and Gentner (2005) showed that presenting children with the labels *top/middle/bottom* improved their judgments in a search task. Researchers hid a sticker on a shelf and children who were provided a relational term (i.e., “See? I’m hiding the sticker on the middle shelf”) to describe the hiding place performed better than children who were instructed with a generic term, like *here*. Casasola (2005) similarly found that priming 18-month-olds – by providing them with a familiar, relational term (*on*) – enabled these infants to categorize instances of support, whereas providing infants with a general verbal command (*Look!*) or no term at all did not aid infants.

Hermer-Vasquez, Moffet, and Munkolm (2001) provide similar evidence indicating that knowing the correct, specific label leads to successful performance in a *right* and *left* reorientation and retrieval task. In their study, preschoolers who could correctly produce the terms *right* and *left* on their own were better at retrieving a hidden object that was to the *right of* or *left of* a prominent feature in a room than children who did not know these terms. Previous work by Hermer and Spelke (Hermer & Spelke, 1994, 1996) suggested that younger children (18-24 months old) could not use a feature as a relational guide which might be construed to suggest the children could not encode the location of a hidden object as *right-of* or *left-of* because they lacked these specific terms in their vocabulary, further suggesting that language acquisition was a key factor. However, because the study was correlational, it is possible that a third factor correlated with language development was responsible for both the acquisition of the terms and improved performance in the orientation task. In a follow-up study, they found that adult performance was disrupted by a concurrent verbal shadowing task that blocked adults’ *accessibility* to labels, strengthening their argument that language was the critical factor (Hermer-Vasquez, Spelke, & Katsnelson, 1999). Taken together, findings from these studies led the experimenters to conclude that labels served to bind together relevant information. Importantly, in both Hermer’s and Gentner’s studies, inaccurate or imprecise language did not improve performance – lending support to the idea that precise, accurate labels for spatial relations are necessary—a question that we also address in the current work. Alternatively, because adults were denied *access* to the appropriate verbal labels, these results also point to the possibility that it is *accessibility* to linguistic coding that drives performance. Taken together, the evidence indicates that language affects the ability to encode or remember, spatial relations and locations, but it is not clear whether having the term in their vocabulary (*acquisition*) is all that is needed or whether *accessibility* of the label – through priming or retrieval – is an additional required step.

Are precise, accurate labels necessary?

Given the existing evidence, it is not clear whether the precise, accurate term is necessary to aid relational judgments or if any label would do. Some studies found that only certain labels aid relational judgments. For example, recall that Loewenstein and Gentner (2005) found that children who were provided the precise, accurate label performed best in a retrieval task. In a similar vein, Shusterman, Lee, and Spelke (2011) found that preschoolers performed better in a reorientation task when they were supplied with linguistic cues pertaining to the precise object location (“I’m hiding it at the red wall”) than when they were provided generic cues (“I’m hiding the sticker over here”) or nonspatial cues (“Look at the pretty red wall”). Similarly, Dessalegn and Landau (2008) – who studied 4-year-olds’ ability to remember a patterned square – found that children had difficulty remembering the specific red-green relation depicted in a square unless they were told the specific relation (e.g., “the red is on (*the left of*) the green”) during the sample presentation - despite their inability to correctly label *right/left* relations outside of this task. Findings from studies like these suggest that the ability to make relational matches is supported by the use of the specific linguistic labels that denote relational concepts (Gentner, 2010; Gentner & Ratterman, 1991). However, Landau’s work (see also Landau & Lakusta, 2009) additionally suggests that the effects of language may be transient, and thus are not always be found, particularly since the children studied by Dessalegn and Landau (2008) could not correctly label “right” or “left”.

Opposing the idea that specific labels are necessary is evidence suggesting that words other than precise, accurate relational terms also promote relational judgments. Christie and Gentner (2013) found that providing 2- and 3-year-olds with a novel noun (e.g., *This is a truffet, can you find another truffet?*) improved their performance on a *same/different* relational match-to-sample task. In this task, children had to match shape relations depicted on cards; when trained on the labels *same* and *different*, only 3-year-olds improved in their performance; however, when a novel noun was applied to the card, both 2- and 3-year-olds made more relational choices. Christie and Gentner (2013) considered this finding as evidence that a novel noun “invites comparison” that “highlights the relational commonalities.” Complementary evidence for the usefulness of other kinds of terms is provided by Son, Smith, Goldstone, and Leslie (2012) in their label-matching study with 4-5-year-olds. In their study, children were able to remember the relation shared by three objects (either a “sharing” or a “pulling” relation) both when novel and known labels were provided. That is, children generalized the named relations to new object sets when provided with novel terms (e.g., using the label Ko-Li-Ko to designate an ABA relation) or known terms (e.g., “pushing” or “pulling”). In short, some studies find that only the correct relational labels aid relational judgment whereas other studies suggest that other kinds of terms also aid performance. If applying *any* label helps relational processing, then it may be that language serves to anchor the relevant information when committing it to memory. In this case, the label might be helping to mark the information for later use and memory retrieval. This is a possibility which we explore in this paper.

Issues with Previous Findings

One reason that it is difficult to draw clear inferences from past studies is because of differences across tasks, relations, and the age groups tested. For example, Loewenstein and Gentner (2005) studied knowledge of vertical relations (e.g., *top*, *middle*, *bottom*) in 4-year-olds, and Hermer-Vasques, Moffit and Munkholm (2001) studied knowledge of horizontal relations (e.g., *right* and *left*) in 6- and 7-year-olds. Loewenstein and Gentner (2005) used a task in which children had to apply a mapping strategy while Hermer-Vasques, Moffit and Munkholm's (2001) task required children to navigate inside a room. Both these tasks required a child to move his or her whole body as part of their response, which may be an added burden to the otherwise cognitive task. Although these studies need not contradict each other, it is more difficult to combine their findings into a cohesive story without evidence from a single study in which knowledge of all the relations are measured with similar verbal and nonverbal tasks. Our current paper provides such evidence; it examines both verbal and nonverbal knowledge of vertical and horizontal relations within a single group of children. In our study, children ranged from 5 to 10 years of age – a broader age than has yet been examined – which allowed us to make direct comparisons across age groups. We expected that verbal knowledge of *above* and *below* would be in place for children of all ages, but that *right* and *left* would be known only by the older children. If language acquisition is the critical driving factor, we should find better *nonverbal* performance on the vertical plane (*above-below*) than on the horizontal plane (*right-left*) in children who have acquired the words *above* and *below* but not *right* and *left*.

Another issue involving previous work is the degree to which flexible relational processing was actually engaged. In many spatial relation judgment tasks use of intrinsic and extrinsic frames of reference have been confounded, with correct performance needing only to apply an extrinsic frame of reference. For example, the children who were successful in Dessalegn and Landau's (2008) tasks could have considered only the extrinsic frame of reference of the rectangles' locations in the square rather than the intrinsic relation between the colored pieces. In this way, children could have simply remembered the *locations* of the red and green rectangles with respect to the screen's frame (extrinsic frame) rather than with respect to the relation of the rectangles to each other (an intrinsic frame of reference). The evidence from infant studies (Gava, Valenza & Turati, 2009; see also Quinn, 2007) is similar in that it could involve making perceptual similarity judgments by attending only to the *side* on which an object appears rather than encoding the relation between two objects (e.g., *to the right of the line*). The current study examines children's ability to apply an intrinsic frame of reference when it matches the extrinsic frame of reference and when it does not.

To examine the ability of children to make flexible relational judgments – judgments that involve focusing only on the intrinsic relation and ignoring the extrinsic frame– children had to encode the relative position of a dot to a line (*above*, *below*, *right*, *left*) that was encompassed by a circle. In this case, the relevant frame of reference was the dot relative to the line within the circle (the intrinsic frame of reference). The circle also appeared at different positions on the screen, which sometimes created conflict between the intrinsic and the extrinsic relation. On “congruent” trials, the intrinsic and extrinsic frames of reference matched; therefore, children could apply either frame to make a correct judgment. For

example, when the dot was below the line and the circle was also on the bottom half of the screen both frames of reference indicated that the correct answer should be *below*. On “incongruent” trials, however, the task required children to abandon the extrinsic frame of reference offered by the screen. For example, when the dot was below the line and the circle appeared on the top half of the screen, the correct answer was also *below*. Thus, in half of the trials children had to overcome incongruencies in the circle's location compared to the screen. Such incongruencies – when two competing frames of reference are available – have been shown to affect both monkeys' (Fortes, Merchant, & Georgopoulos, 2004) and young children's (Smith, Ratterman, & Sera, 1988) performance. For example, both find it more difficult to judge which of two items is *higher* when the items differing in height appear at the bottom of a board (Smith, Ratterman, & Sera, 1988) or a computer screen (Fortes, Merchant, & Georgopoulos, 2004). Examining performance across congruent (i.e., when both frames of reference match) and incongruent trials (i.e., when the correct judgment relies only on the intrinsic frame of reference offered by the dot and the line) enables us to examine the ability to make relational judgments, as has been often done in past work (on congruent trials), as well as those that require more flexible relational coding (on incongruent trials).

A final issue that we addressed involves how we manipulated language accessibility. In much past work (e.g., Lowenstein & Gentner, 2005; Dessalagn & Landau, 2008), the precise, accurate relational term was provided by the experimenters. It is not clear whether children can spontaneously *access* and *use* the specific, correct terms when they are not explicitly cued to do so. In our study, instead of explicitly training the child to use certain terms or providing a to-be-used label, we let children use labels that they generated themselves by having them perform a verbal task before performing a nonverbal task (i.e., so they *activated* their own labels). Furthermore, in this way, precise, accurate labels were not necessarily used (as measured by performance on the verbal task). This strategy also enabled us to address the related issue of the nature of the labels used by the child. Recall that in past work, different types of labels (e.g., relational vs novel nouns) led to improved performance suggesting that precise labels may not be necessary to enhance the ability to make nonverbal relational judgments. By having children generate labels themselves, we, thus, also examined whether generating the precise, accurate label improved performance as much as another (even incorrect) term.

This Study

We examined how verbal and nonverbal knowledge of vertical and horizontal spatial relations (*above*, *below*, *right* and *left*) develop in children. In four experiments, we compared within group differences on these judgments. Relational knowledge has often been shown to develop extensively from ages 4 to 5 years (Gentner, Anggoro, & Klibanoff, 2011); therefore, our chosen age range of 5- to 10-year-olds should include children who have at least some relational knowledge. In the first experiment, we examined our broad age range of 5-10-year-olds and used the results from this experiment to identify a target age group on which to focus subsequent experiments. Differences in language knowledge within the same group of children constitute a naturally-occurring “manipulation” of language that should be independent of more general cognitive skills and is ideal for comparing verbal

performance to nonverbal performance. If language systematically precedes and causes the ability of nonverbal relational judgments to improve, then we should find better *nonverbal* performance along the vertical (*above-below*) plane than the horizontal (*right-left*) plane.

Our predictions were as follows. We predicted that if *acquisition* by itself was responsible for flexible relational coding, then performance on the nonverbal tasks should parallel performance on the verbal tasks in Experiments 1 and 2: children who know precise, accurate labels (e.g., *above and below*) should also do well on the corresponding nonverbal instantiations of those relations. We used a task similar to the one used by Goodwin and colleagues' (2012) studies with monkeys which found better performance on *right-left* (horizontal) judgments than on *above-below* (vertical) judgments. The Goodwin et al. (2012) findings contradict some evidence from humans suggesting better performance on the vertical than on the horizontal plane (e.g., Clark, 1972; Hayward & Tarr, 1995; Landau & Hoffman, 2005; Carlson, West, Taylor, & Herndon, 2002): this study will, also offer evidence relevant to the potential "privilege" of the vertical over the horizontal plane. Experiments 3 and 4 explored the roles of accessibility and practice.

In Experiment 1, we measured verbal knowledge of the four relations using a static task. Then in Experiment 2, to address the possibility that the static verbal task did not require the same kind of flexibility required by the nonverbal task, we modified the verbal task used in Experiment 1 to a dynamic one, which included both intrinsic and extrinsic frames of reference, and paralleled the nonverbal task. In Experiment 3, we predicted that increased *accessibility* to verbal codes would affect subsequent performance on the nonverbal task. Importantly, we examined children's self-generated use of language instead of providing them with particular labels as has been done in previous studies. This distinction is an important addition to the literature because children do not receive overt guidance in their everyday lives and understanding how children come to use labels naturally has real-world implications. Finally, in Experiment 4 we examined the potential role of practice on performance.

Experiment 1

The goal of this experiment was to examine the relation between the ability to encode and remember the relations *above*, *below*, *right* and *left* verbally and nonverbally in 5-10-year-olds. Previous studies have suggested a strong relation between verbal encoding of spatial relations and performance on nonverbal spatial memory tasks. We expected to find differential nonverbal performance if such performance is enhanced after the corresponding terms have been acquired. So, if language causes an immediate improvement in the ability to extract and remember relations nonverbally, then we should find better performance for *above* and *below* than for *right* and *left* in our nonverbal task since knowledge of the words *above* and *below* typically precedes knowledge of the words *right* and *left*. In order to avoid potentially priming children to apply a verbal strategy in the nonverbal task, all children completed the nonverbal task first. Our verbal task was modified from Cox and Richardson's (1985) paradigm, while our nonverbal task was new and based on the paradigm used by Goodwin et al. (2012) in their studies with monkeys.

Method

The methods have been published previously (in Scott et al., 2015), but we reiterate them here for convenience. The analyses from our previously published study (which were completely different from the current ones) showed that performance across a range of tasks of relational knowledge becomes more similar and efficient with age. However, there was no specific focus on the relation between verbal and nonverbal performance in our previously published work.

Participants

A total of 120 monolingual English-speaking children from 5;0 to 10;11 years of age from the Minneapolis and St. Paul metropolitan areas participated in the study. The children were contacted from a list of interested families and had no known language or cognitive deficits. Equal numbers of girls and boys participated and they performed equivalently. Table 1 contains descriptive demographic information. Almost all of the children were Caucasian and from middle to high SES families. An additional 13 children were tested in a pilot study, but their data were not included.

Nonverbal task

This task required the child to encode and remember whether a dot had appeared to the *right*, *left*, *above* or *below* a line within a circle (Figure 1a). The dot and line were inside a circle that could appear in one of four quadrants on a computer screen (Figure 1b). Each relation appeared in each of these quadrants 3 times (and in three different perimeter positions within each quadrant) for a total of 48 trials. The order of presentation was randomized for each child. Children made their responses by touching either side of a divided screen: the blue or red side when the screen was divided vertically (*above-below* trials) or the yellow or green side when the screen was divided horizontally (*right-left* trials; Figure 2). Trials consisting of the dot being *above* the line within the circle were congruent if the stimulus appeared in quadrants 1 or 2, but were incongruent if they appeared in quadrants 3 or 4. The opposite was true for *below* stimuli. Similarly, trials in which the stimulus for *right* was congruent were when the stimulus appeared in quadrants 2 or 3, and incongruent when it appeared in quadrants 1 or 4. The opposite was true for *left* stimuli.

Feedback was provided in the form of a 3-second audio clip of applause after a correct answer, and a 1-second zapping audio clip after an incorrect response. Participants had 10 seconds to complete each trial. If the participant did not indicate a choice by touching the screen within 10 seconds, the trial ended and was recorded as an error; this happened in less than 2% of all trials (across all experiments). A black screen appeared between each trial for 1 second. Children performed this task first, followed by the verbal task.

Before testing began, children were trained with 8 trials: one trial of each relation in each congruency type. All children started with the same set of eight training trials before moving on to the test trials. To avoid the use of the relational terms under investigation, the experimenter only said “here” and “there” and never used the terms “above”, “below”, “right” or “left” in providing instructions on how to play the nonverbal computer game.

During training, the experimenter showed the child how to play the game by performing the first three trials of the game correctly but purposely answered the fourth trial incorrectly to show what happened when an incorrect answer was provided, so that the child would not be startled by the zapping noise during the testing phase or intentionally respond incorrectly to see what would happen. When the eight training trials were finished, the experimenter asked, “Do you think you know how to play the game?” If the child said “yes” then the test trials were started and the child was no longer provided with any instructions from the experimenter. No child needed additional training.

Incongruent and Congruent Trials—The relational position of the dot inside the circle was either congruent or incongruent with respect to the location of the circle on the screen. The task was dynamic because the relational judgments varied across different positions on the screen from trial to trial. Trials where the relation between the circle and the screen matched were congruent, while trials where these relations did not match, were incongruent (see Figure 2).

Verbal task

We tested children's ability to produce and comprehend the labels *above*, *below*, *right* and *left* using a task similar to the one used by Cox and Richardson (1985), which has been used previously to test the ability to produce relational terms in other studies (e.g., Hermer-Vazquez, Moffet, & Munkholm, 2001). Our version of the apparatus consisted of two magnetic dry-erase boards with black lines drawn to resemble a “Tic Tac Toe” board (see Figure 3) and four magnets; it is described in more detail in Scott et al., (2015). The order of presentation of the magnets was consistent across both verbal tasks and as follows: cat, dog, house, and child. The relative location of the magnet for each trial was randomized using a Latin square design. In this design, each magnet was seen in each relation only once and each relation was seen only once per four trials making 16 trials total in each task. Neither training nor feedback was provided. Children performed the production task first to avoid cueing children to the correct answers on that portion of the verbal task, as these answers were provided in the comprehension task.

Production Task—In the production task, children were asked to watch the experimenter place one magnet on the board and then to describe the location of the magnet with respect to a centralized circle to another person who could not see the board using just one word. Sometimes children offered only a vague term, such as “here” or they pointed to the board in lieu of producing any term. In these cases, children were encouraged to “think of a different word that would help the person who couldn't see [their] board understand where the [magnet] was”, since they were performing the task under the ruse that the other person was trying to make their board look like the child's board. Children were not told the nonspecific word was incorrect. Children's answers were considered correct if they produced the words *above*, *below*, *right* or *left*, or a closely related but equally specific term such as “on top” for *above* following Cox & Richardson (1985) since we were more interested in children's knowledge of the vertical and horizontal spatial relations than on these rather specific and often interchangeable terms. “Side” was not accepted as a correct response for *right* or *left*,

but children were encouraged to think of a different word that was more specific since “side” is an otherwise accurate term.

Comprehension Task—In the comprehension task, children were asked to place a magnet *above*, *below*, *to the left of* or *to the right of* the circle on the board. In this task, there were 3 (out of 8) locations that were counted as correct: directly above the circle and the two adjacent corners (e.g., top-right corner or top-left corner for “above”). Only two children utilized the corners of the board. Chance was set at 37.5% (3 out of 8 possible locations) for the comprehension task, but there was no chance level set for the production task since children could produce any number of terms.

Results

We were interested in two questions: (1) whether performance was consistently better on *above* and *below* relations than *right* and *left* relations in both tasks; and (2) whether individual performance on the verbal task was correlated with performance on the nonverbal task. For the verbal tasks, we compared performance on the two relational planes (i.e., *above* and *below* as the vertical plane, *right* and *left* as the horizontal plane), rather than on each relation, since children have been shown to learn relational pairs together, first learning that they are opposites then learning the correct spatial mapping (Clark, 1972; Scott et al., 2015); we compare performance on each relation separately for the nonverbal task. For each analysis, we used nonparametric tests since performances were highly skewed and bimodal. We examined correct performance for each age group separately so that we could map the developmental trajectory of relational processing. We also used these results as a basis for focusing on the most relevant age group in subsequent experiments.

Verbal tasks

For the production task, a Wilcoxon Signed Rank test comparing performance on the two planes indicated that performance was better for *above-below* judgments than for *right-left* judgments in 5-year-olds ($z = -3.5, p < .01$), 6-year-olds ($z = -2.6, p < .01$), and 7-year-olds ($z = -2.0, p < .05$), but not significantly different for 8-, 9- or 10-year-olds. For the comprehension task, a Wilcoxon Signed Rank test indicated better performance on *above-below* judgments than *right-left* judgments in 5-year-olds ($z = -2.7, p < .01$), but not for any other age group. The data are plotted in Figure 4, which illustrates the steady increase in performance with increasing age in both verbal tasks with children attaining ceiling performance around age 8 years.

It is clear from Figure 4 that most development is occurring on the terms *right* and *left*; performance on *above* and *below* appear to be at or near ceiling in all six age groups. It is also clear from Figure 4 that production and comprehension follow the same developmental patterns; however, and as has been found in other domains, the ability to correctly comprehend terms develops earlier than the ability to correctly produce these terms (e.g., Benedict, 1979; Clark & Hecht, 1983). Moreover, within the age groups studied, it appears that the most development is occurring on the ability to produce the terms, especially *right* and *left*, while comprehension appears much closer to ceiling from the youngest age. Since we were interested in comparing performances of children who knew the correct relational

terms to children who did not know the correct terms on nonverbal performance, we focused on the task that elicited the most variability in performance in subsequent analyses: verbal production. When analyzing the relation between verbal and nonverbal performance in subsequent analyses we operationalized verbal performance as performance on the production task. Another reason for examining production (vs. comprehension) is that several studies have found that production (and not comprehension) is most closely related to nonverbal performance (e.g., Hermer-Vasquez et al, 2001).

Nonverbal task

We used the results from the verbal production task to divide children into 3 age groups: the age preceding knowledge of the relational terms (5 years), the age with partial knowledge of the relational terms (6-7 years), and the age with complete knowledge of the relational terms (8-10 years). Figure 5 illustrates correct nonverbal performance by the three age groups. To further examine the first question, we conducted a Friedman test to determine whether the distribution of performance differed between relations and across congruency types. We chose to analyze each relation separately rather than by plane to avoid making any assumptions about nonverbal performance on the different relations *a priori*.

Results indicated that 5-year-olds performed reliably better on congruent than incongruent *above*, *below*, *right* and *left* trials ($\chi^2(7) = 42.6, p < .001$). The associated pairwise comparisons with unadjusted significance levels indicated better performance on all congruent trials than all incongruent trials both within and across relations. We did not adjust for multiple comparisons because these were planned comparisons. For 6- and 7-year-olds, results showed that overall, they also performed reliably differently on congruent and incongruent *above*, *below*, *right* and *left* trials ($\chi^2(7) = 18.4, p < .01$). The associated pairwise comparisons indicated better performance on most congruent trials than on incongruent right trials (*above*: $z = 1.9, p = .052$; *below*: $z = 2.7, p < .01$; *right*: $z = 2.3, p < .05$; *left*: $z = -2.2, p < .05$). It is interesting that *above* and *below* incongruent trials are tightly associated, and performed slightly more accurately (though not significantly so) than *right* and *left* incongruent trials, which are also tightly associated with each other. For 8- to 10-year-olds, the distributions of congruent and incongruent *above*, *below*, *right* and *left* trials were not significantly different.

It is clear from Figure 5 that the most development on the nonverbal task is occurring on the incongruent trials across the age groups while performance on congruent trials appears to be at ceiling for all groups. Because congruent trials can be solved by either intrinsic or extrinsic relational codes, and incongruent trials require focusing only on the intrinsic relation and ignoring the extrinsic one, we operationalize the ability to make *flexible* nonverbal relational judgments as performance on the incongruent trials in subsequent analyses. Furthermore, because *above* and *below* are tightly coupled and *right* and *left* are tightly coupled, we also combined the relations within each relational plane in subsequent comparisons.

Overall, we did not find that nonverbal performance on *above-below* trials was reliably better than performance on *right-left* trials as predicted by the hypothesis that language, immediately after being acquired, automatically causes improvement in the ability to make

nonverbal relational judgments. This hypothesis suggests that nonverbal performance should parallel verbal performance, but we did not find evidence that it does. However, it is possible that the large individual variation that we observed obscured the relation. Thus, we examined individual performance more closely and directly compared each child's verbal performance to their nonverbal performance via correlational analyses.

Comparison across verbal and nonverbal tasks

It might be argued that individual children may be using a linguistic strategy to perform the nonverbal task, if language is accessed automatically. For instance, when children see the dot as being to the “right” of the line a linguistic representation of “right” may be instantly activated (as has been argued to occur in Stroop tasks; Diamond, Kirkham & Amso, 2002; Dalrymple-Alford, 1973; Klein, 1964). If children were attaching a correct, precise verbal label to the dot-line relation, then we would expect to find a predictive relation between verbal and nonverbal performance. This possibility was addressed directly by our second question, which asked whether knowing the term for the relations facilitated successful nonverbal performance. To address this question, we examined whether performance by individual children on the verbal task was correlated with their performance on the nonverbal tasks. We compared only the incongruent trials of the nonverbal task to verbal production since the incongruent trials were the most difficult for children at the same time their ability to produce the terms was still developing. We looked at each age group separately since performance varied by age.

Performance was not normally distributed, rather it was bimodal with children either doing quite well (near ceiling) or quite poorly (near floor performance), so we used Spearman's Rho correlation test on the raw percentage scores. Only correlations that matched relations across tasks are relevant to our hypotheses since knowing the relational term is expected to improve performance for the corresponding relation; however, we show all correlations in Table 2. No correlations were reliable among the 5-year-olds. Among the 6-7-year-olds, we found only the unpredicted correlation between production of *left* and performance on incongruent *below* trials. Among the 8-year-olds, producing the term *left* reliably predicted nonverbal performance on *left* incongruent trials, and we found significant (unpredicted) correlations between production of *right* and incongruent below trials, and production of *left* and the incongruent below trials. Thus, we found only one (out of a possible 12) reliable correlations of the predicted ones. Because the only correlations are with production of *right-left*, it could be that production of these terms facilitates flexible relational coding for all four spatial relations. However, the number of correlations that support this hypothesis are also small, 3 out of 24. The fact that the only reliable correlations between verbal and nonverbal performance were found among the older children also points to a period of time passing between acquisition of the terms (in 5-year-olds) and their possible impact.

Analyses with a Loess Approach

Despite not finding significant relationships between verbal and nonverbal performance by individuals, we did find that verbal knowledge develops slightly before nonverbal knowledge. Figure 6 displays the developmental trajectory of both verbal and nonverbal knowledge of the four spatial relations tested. This figure shows the average performance of

children using a sliding window approach (i.e., a Loess graph). This approach calculates the average performance of the youngest 24 children in the sample (i.e., $n_1:n_{24}$) and represents that average as a filled circle for incongruent nonverbal trials (or an unfilled triangle for verbal production), then the window slides over to include the next older child and calculates an average performance for that child and the 23 children immediately younger (i.e., $n_2:n_{25}$). This approach smooths out individual differences in performance and finds the average for the parameterized age group, thus making the data more smooth and continuous. It thus offers a more fine-grained picture of development on the relevant tasks while ignoring individual differences. The pattern of performance illustrated in these plots indicates that verbal ability does appear to lead nonverbal ability but only slightly in the case of *right* and *left*. These Loess curves indicate that in the *above/below* relational plane, production of the labels fully precedes nonverbal incongruent abilities; however, in the *right/left* relational plane verbal production increases alongside success on incongruent trials until about the age of 80 months, at which point verbal production reaches ceiling performance but ceiling performance on incongruent trials has not yet been attained. This observation of slightly earlier correct verbal production of *right/left* than the ability to overcome nonverbal incongruency, as a population measure, could be due to acquisition of the words, or it may be tied to a third, unknown factor. Experiment 3 addresses this possibility.

Discussion

With respect to verbal performance, we found an advantage of the words *above-below* over the words *right-left* for 5-7-year-olds for both production and comprehension, but no difference for 8-10-year-olds who performed at ceiling on both verbal tasks. We did not, however, find analogously consistent and reliable differences in the ability to make nonverbal relational judgments in the *above-below* (vertical) relational plane over the *right-left* (horizontal) relational plane. If learning language, immediately after acquisition, automatically improves the ability to make nonverbal relational judgments, then we should have found better performance on nonverbal judgments of *above* and *below* compared to *right* and *left*. Converging findings on the relation between verbal and nonverbal performance emerged from our correlational analyses indicating that verbal performance was not systematically related to nonverbal performance, overall. Thus, these findings do not offer strong support for the idea that acquisition of relational terms by itself consistently and automatically improves nonverbal performance. However, our results also consistently show (though not significantly so) that verbal performance precedes the ability to focus (and ignore) different spatial frames of reference.

One possible reason why verbal performance preceded but was not consistently shown to facilitate nonverbal performance was that our static verbal task was not measuring the same ability to make flexible relational judgments as our dynamic nonverbal task. In our nonverbal task, every relation of *above-below/right-left* appeared in a different position on the screen on every trial, thus requiring children to impose and abandon different frames of reference on each trial. In contrast, our verbal task (which has been used by other researchers) only required that children impose one extrinsic reference point – the middle of the tic-tac-toe board – that could have been used on trial after trial. Perhaps it is the ability to dynamically impose and abandon relational standards, which language use typically

requires, that drives the ability to make dynamic relational judgments. We addressed this possibility in Experiment 2 with a new verbal task. Again, we only consider verbal production performance in the remaining experiments because the correct verbal production of the terms was still developing while comprehension was relatively at ceiling for most children. We also focused on 6-7-year-olds in the following experiments because children at this age are still learning *right-left* relations and thus their knowledge reflects a range of knowing and not knowing these terms.

Experiment 2

The results from Experiment 1 did not clearly support the idea that acquisition of relational terms immediately and automatically improves the ability to make nonverbal relational judgments. However, the pattern of findings could have been due to the fact that ignoring an extrinsic frame of reference was required in the nonverbal task, but not the verbal task. The relational judgments required in the verbal task of Experiment 1 might have been easier than the ones required in the nonverbal task. So, we tested children's verbal knowledge using a task that required dynamic relational judgments. If the strategic tool that language offers is a medium for adopting and abandoning (i.e., shifting) relational standards, then we might find a systematic pattern of results across both relational planes when both verbal and nonverbal tasks require the same dynamic and flexible relational skills. We use our results from Experiment 1 to focus on verbal production, and on 6-7-year-olds.

Participants

Children from 6;0 to 7;6 years of age were recruited from the same population that participated in Experiment 1, but none had participated in the previous experiment. Twelve boys ($M = 80.3$ months, $SD = 5.44$) and thirteen girls ($M = 80.5$ months, $SD = 5.51$) participated for a total of 25 children. An additional 2 children were tested but their data were not used because of an error in the computerized presentation of the stimuli. We chose to limit the age of 7-year-olds to 7;6 since children these ages are more likely to still be learning *right-left* terms while having solid knowledge of *above* and *below*. For our comparison across experiments, then, we only used the performance of the 30 children from Experiment 1 who matched this age range.

Nonverbal relational judgment task

We used the same nonverbal task that we used in Experiment 1 and, as in that experiment, children performed the nonverbal task first.

Dynamic verbal task

For this experiment, the verbal production task consisted of a modified version of the nonverbal computer task (see Figure 7). However, instead of touching the screen to indicate a relational judgment as they did in the nonverbal task, the child was asked to verbally label the relation depicted inside the circle. Thus, the final choice screen was deleted and the visual static screen was extended from 3 seconds to 9 seconds, for a total trial length of 10 seconds. Just as in the nonverbal task of Experiment 1, the reference object appeared in each of the four quadrants in a random order, thus correct performance on this task required

flexible and dynamic verbal relational coding. As in Experiment 1, children's answers were considered correct if they produced the words *above*, *below*, *right* or *left*, or a closely related but equally specific term such as “on top” for *above* following Cox and Richardson (1985) since we were more interested in children's knowledge of the vertical and horizontal spatial relations than on these rather specific and often interchangeable terms. “Side” was not accepted as a correct response for *right* or *left*,

Unlike the previous, static verbal task, children were given up to 8 training trials to ensure they understood the task; however, they received no feedback. If a child produced a nonspecific word during training, like “side” for *left* or *right* relations, then they were encouraged to “think of a different word that would help someone who couldn't see the screen understand where the dot was”, but they were not told that the nonspecific word (e.g., “side”) nor any label was incorrect. At the end of training, children were asked, “Do you think you know how to play the game?” If the child said “yes” then the test trials were started. Again, children received no feedback on correctness during the test phase nor were they asked to produce a more specific word during testing since a time limit was now imposed on responses. They completed a total of 48 dynamic verbal trials.

Results

We were interested in two questions. One was whether we would replicate our results of Experiment 1 with a verbal task that required more dynamic relational judgments. If we found the same pattern of performance on the verbal and nonverbal tasks as in Experiment 1, then we would be more confident of the relation between verbal and nonverbal performance that we found in Experiment 1. A related question was how performance on the dynamic verbal task would correlate with nonverbal performance. Finally, we also directly compared overall performance between Experiments 1 and 2.

Dynamic verbal task

As in Experiment 1, children's performance was bimodal in the verbal task. For the verbal production task, a Wilcoxon Signed Rank test comparing overall performance on the two planes indicated that performance was better for *above-below* judgments than for *right-left* judgments in (Figure 8; $z = -3.1$, $p < .01$). Performance is plotted in Figure 8. However, since the dynamic verbal task included both congruent and incongruent trials, we conducted a Friedman's test to determine whether the distribution of performance differed across congruency types. Results indicated, overall, that distributions were significantly different for *above-below* versus *right-left* trials ($\chi^2(3) = 25.2$, $p < .01$). Contrary to our findings from the nonverbal task in Experiment 1, however, we did not find differences in distribution between congruent and incongruent trials in our dynamic verbal task, but rather only between the vertical and horizontal planes. Children performed better on incongruent *above-below* trials than on both congruent ($z = -2.2$, $p < .05$) and incongruent *right-left* trials ($z = 3.0$, $p < .01$). They performed better on congruent *above-below* trials than on incongruent *right-left* trials ($z = 2.6$, $p < .01$) and marginally better than on congruent *right-left* trials ($z = 1.9$, $p = .06$). Overall, we replicated our findings in Experiment 1 for the verbal task; namely, of a different pattern of performance across tasks within the *above-below* relational plane

compared to the *right-left* relational plane, but using a dynamic verbal production task. Importantly, we found no effect of congruency in the verbal task of this experiment.

Comparison between verbal and nonverbal tasks

As in Experiment 1, we investigated the relation between verbal and nonverbal performance. Because we have both congruent and incongruent trials in the modified verbal production task, we analyzed these separately; although, it may be expected that verbal incongruent trials would mostly mirror nonverbal incongruent trials. Our hypothesis was that if language knowledge automatically elevates nonverbal performance, then verbal performance should have a reliable relationship to (i.e., predict) nonverbal performance, particularly on incongruent trials. Nevertheless, we use a two-tailed test to examine for the opposite possibility: that nonverbal performance drives verbal performance. As in Experiment 1, a Spearman's Rho correlation did not indicate a reliable relationship between verbal and nonverbal performance. Table 3 shows the correlation table.

Comparison across experiments

The impetus for this experiment was to make sure that our verbal task required the same degree of relational coding as our nonverbal task, and we generally replicated our results from Experiment 1. To directly confirm this finding, we compared performance on the verbal tasks across the experiments directly. For comparison purposes, we combined over congruent and incongruent trials of the dynamic verbal task in Experiment 2 to get a total performance on each relational plane. A Kolmogorov-Smirnov test on the distributions yielded no significant differences between the results for Experiment 1 and Experiment 2 (see Figure 9).

With regards to performance on the nonverbal tasks, we expected no difference in performance since in both experiments the nonverbal task was performed first, and therefore, not affected by the nature of the subsequent verbal task. A Kolmogorov-Smirnov test on the distributions to check that the samples were from the same populations indicated no significant differences, so we considered the two experimental groups to be equivalent.

Discussion

After changing the verbal task to include the ability to ignore the extrinsic frame of reference, our analyses indicated that this was not a factor in children's ability to name the relations, even incongruent ones. Additionally, despite the greater computation and attention required to succeed in the dynamic verbal task (as compared to the static verbal task used in Experiment 1), performance on these two verbal tasks were not significantly different from each other. One new finding was that congruency did not affect verbal performance, despite it having a significant effect on nonverbal performance. This finding suggests that children overcome incongruency earlier in a verbal task than a nonverbal one.

As in Experiment 1, we tested for a reliable relationship between performance on the verbal production task and the nonverbal incongruent trials and again found none. In other words, we did not find that knowledge of the words systematically co-varied with the ability to make nonverbal judgments. Overall, these results, again, do not strongly support the idea

that knowledge of relational terms by itself (i.e., *acquisition*) consistently and *automatically* improves the ability to make nonverbal dynamic relational judgments across different relational planes in spite of the fact that the ability to make relational judgments verbally precedes the ability to make these judgments in a nonverbal task. So, the element of temporal precedence between linguistic and nonlinguistic knowledge has been established towards meeting the criterion for a causal relation between language and cognition. However, we did not observe the critical second element - unique co-variation. So it is possible that a third factor, unrelated to language knowledge or language use, is causing the improvement in the ability to make nonverbal incongruent judgments. In Experiment 3, we attempted to naturally activate the category labels during the nonverbal task by having children perform the verbal task first. If a third variable, unrelated to language knowledge or use, is responsible for causing improvement in nonverbal judgments, this linguistic manipulation in Experiment 3 should not affect nonverbal performance.

Experiment 3

Evidence indicates that knowing labels (i.e., having labels available) and accessing them are separate abilities (see Brod, Werkle-Bergner, & Shing, 2013 for similar argument of distinguishing between availability and accessibility). If the role of language in promoting relational judgments is to offer a strategy for ignoring irrelevant frames of reference, then we might find better performance on a nonverbal task when category labels are made more accessible. Kranjec, Lupyan and Chatterjee (2014) used a similar strategy when studying the role of language in spatial judgments. To increase accessibility, we had children perform the verbal task before the nonverbal task and examined whether nonverbal performance improved (i.e., became less variable) in comparison to Experiments 1 and 2 on the incongruent trials. If relational language supports nonverbal performance, activating linguistic knowledge should improve children's performance on the nonverbal task. If a third variable (unrelated to language knowledge or use) is responsible for improved nonverbal performance this manipulation should not affect performance.

One potential problem of having children perform the verbal task first is that children might not use the relational terms intended (*above*, *below*, *right* or *left*); however, evidence from some studies indicates that using other terms may nonetheless improve performance on a nonverbal task (e.g., novel nouns: Christie & Getner; 2013). Thus, we also assess the degree to which use of precise labels influenced performance by analyzing performance on the nonverbal task as a function of performance on the verbal task. If specifically coding the dot as *to the side* of the line is equivalent to coding it as *to the right* of the line then, we should see no difference in performance on the nonverbal task as a function of the specific words used in the verbal task. Moreover, if using *any* term is sufficient, then we should find improvement in the nonverbal task, regardless of the precise label used. Nonverbal performance should not improve if language does not influence it.

Methods

Participants—Children from 6;0 to 7;6 years of age were recruited from the same population that participated in Experiments 1 and 2, but none had participated in either of

the first two experiments. A total of 25 monolingual, native English-speaking children with no known cognitive differences were tested. Eleven boys ($M = 81.5$ months, $SD = 6.36$) and fourteen girls ($M = 82.4$ months, $SD = 3.57$) were tested. An additional 3 children were tested but their data were not used; 1 due to parental interference and 2 due to computer failures.

Nonverbal relational judgment task

We used the same nonverbal task that we used in Experiments 1 and 2, except in this experiment the task was performed second.

Verbal relational judgment task

We used the same verbal task that we used in Experiment 2. The only difference was that performance of this task preceded the nonverbal task. It is important to note that, as in Experiment 1 and 2, no feedback was provided during the testing phase of this task.

Results

We were interested in two questions. The main question was whether we would find improved performance on the incongruent trials of the nonverbal task as compared to the previous two experiments. A related question was whether performance would vary systematically on the nonverbal task as a function of performance on the verbal task. With regard to the second hypothesis, if the effect is from using *any* label rather than a precise one then, we expect to find no correlation between the tasks.

Nonverbal task

We conducted a Friedman's test to determine whether the distribution of performance on the nonverbal task differed between relational planes and congruency types. Results indicated that, overall, distributions were not significantly different for congruent and incongruent *above-below* and *right-left* trials ($\chi^2(3) = 1.6$, $p = .66$). Figure 10 shows the means and standard errors. Thus, in contrast to performance in Experiments 1 and 2 - where children performed better on nonverbal congruent than incongruent trials - performance on nonverbal incongruent trials was as good as performance on nonverbal congruent trials in this experiment. It appears that by performing the verbal task first, children were able to overcome their difficulty with nonverbal incongruent trials in this experiment.

Comparison across experiments

To directly examine differences in patterns of performance across the three experiments we first compared the variance in performances using Levene's test for equality of variances. The most important comparison involved performance on the incongruent trials (Figure 11). If nonverbal performance was enhanced in Experiment 3 by virtue of performing the verbal task first, then we should observe better, more consistent overall performance on incongruent trials of Experiment 3 than Experiments 1 and 2 combined. Indeed, overall ($F_{54,24} = 12.0$, $p < .001$), there was significantly less variable performance on the nonverbal incongruent trials in Experiment 3 than in Experiments 1 and 2 (*above-below*: $F_{54,24} = 15.3$, $p < .01$; *right-left*: $F_{54,24} = 8.0$, $p < .01$). Further evidence is provided in a comparison of the

distribution of the nonverbal incongruent trials of Experiments 1 and 2 versus Experiment 3 using a Mann Whitney U test ($p < .05$). Because the distribution of performance scores was bimodal in Experiments 1 and 2, results from parametric tests could be misleading. However, because much prior work uses parametric tests (e.g., Christie & Gentner, 2013; Dessalegn & Landau, 2008), we include the results from one here so the reader can align our results with previous ones. A comparison between nonverbal incongruent trials of Experiments 1 and 2 and nonverbal incongruent trials of Experiment 3 yields a reliable difference between the means ($t(df = 68) = -2.763$; $p < .01$; equal variances not assumed; mean difference = 15.14; $SE = 5.48$, Cohen's $d = .513$). So, having children perform the verbal task first in this experiment improved their performance on the incongruent trials nonverbal task in comparison to Experiments 1 and 2.

Label Precision

The finding that children performed better overall on the incongruent trials in Experiment 3 than in Experiments 1 and 2 does not identify the specificity of the language effect. Did coding a relation with any label (e.g., *to the side*) improve performance? Or did children need to code the location of the dot with the specific label (e.g., *to the right*) to improve their performance? In order to address this question, we examined performance on all of the tasks for each relation. Thus, as in Experiment 1, we asked, for example, whether correct verbal production of *above* was related to correct performance on the nonverbal trials depicting *above* relations. If knowing the precise relational term, and using it correctly, was responsible for improved performance on the nonverbal task from Experiments 1 and 2 to Experiment 3, then we should find a positive correlation between verbal production and performance on the nonverbal incongruent trials. In short, children who accurately produced the terms *right-left* (or *above-below*) in the verbal task should have done better on these relations in the nonverbal task than children who used other terms.

As before, only correlations that matched relations across tasks are relevant to our hypotheses since knowing the relational term is expected to improve performance for the corresponding relation; however, again, for transparency we show all correlations in Table 4. A Spearman's Rho correlation on the raw percentage correct performances indicated that only one relevant correlation of verbal production to the nonverbal incongruent trials: production of incongruent *right* was correlated with nonverbal incongruent *right* trials.

To further examine whether children who performed better on the nonverbal task had partial knowledge of *right* and *left*, we compared the nonverbal performance of children who mixed up *right* and *left* (who called *left* “right” and vice versa; $n = 11$) with children who produced the generic terms (*side* or *other side*; $n = 7$). We found no differences in nonverbal performance between these two groups using either parametric (independent t-tests) or nonparametric tests (Mann-Whitney U). Thus, we found little evidence that precise application of the relational terms was responsible for improved nonverbal performance. Our findings suggest that the improvement we found in the incongruent trials was due to, for the most part, increased accessibility to labels, even imprecise ones.

Discussion

We found that asking children to produce the relational terms before they performed the nonverbal task led them to make more successful nonverbal relational judgments, especially on the incongruent trials. Thus, although the role of language *acquisition* seems not to be that of immediately improving the ability to make flexible judgments, its role seems to be that of providing an efficient strategy for encoding and remembering relational information. Apparently, acquisition of the labels by itself does not lead to improved nonverbal performance. Rather, it is only when a label is utilized that it simplifies relational judgments, possibly by allowing children to ignore irrelevant details of the stimuli. If children use category labels for “chunking” all instantiations of a relation into a single category (i.e., all *above*), regardless of its location on the screen (i.e., congruency type), then children's performance converges to being similar across different tasks and conditions, and the difference between performance on congruent and incongruent trials disappears. We believe that when children performed the verbal task first they became accustomed to applying linguistic labels to the intrinsic frame and used this same strategy to encode and remember the relation between the dot and the line on the incongruent trials as well as the congruent trials of the nonverbal task. Furthermore, we found that even if the information contained in the verbal codes was not accurate – based on verbal production performance – that the use of any verbal code was enough to increase performance. These findings suggest that children of this age can use language as a strategy for solving problems in an otherwise nonlinguistic task, but may typically fail to use these strategies spontaneously. It seems, then, that language is not accessed *automatically* once terms have been acquired, but rather that these must be *accessed* before they can be applied to an otherwise nonverbal task. Failure to find the use of the specific correct terms suggests that language may be serving a role of “chunking” or as some other memory tag; an explanation that we have suggested in previous work (Scott et al., 2015).

However, a final concern that arises is the degree to which our findings from this experiment reflect a practice effect. Children who performed the verbal task before the nonverbal task in this experiment saw the relations and touched the response screen 48 times before they performed the nonverbal task. Merely seeing the relations and practicing touching the screen may have been the sole drivers of improved performance. To address this concern, we conducted a final experiment in which children performed a nonverbal version of the dynamic verbal task before the experimental nonverbal task (that was used in Experiments 1, 2, and 3). If children perform as well on the nonverbal task in Experiment 4 as in Experiment 3, then the improved performance in Experiment 3 may be explained by solely by a practice effect, not a language accessibility effect.

Experiment 4

If the improved performance on the incongruent trials of the nonverbal task can be attributed solely to having more experience with the relations and practice with responding (i.e., screen-touching), then we would expect to find equally improved performance on those trials regardless of what relational judgment task was performed first. So, in Experiment 4, we gave children a modified, nonverbal version of the dynamic verbal task that had been

performed by children in Experiments 2 and 3, thus taking away only the verbal priming of Experiment 3. We kept all other aspects of the task (i.e., no feedback and immediate response) the same as in previous experiments, which served to differentiate it from the “regular” nonverbal task that followed.

Participants

Children from 6;0 to 7;6 years of age were recruited from the same population that participated in Experiments 1, 2 and 3, but none had participated in any of those previous experiments. A total of 25 monolingual, native English-speaking children with no known cognitive deficits were tested. Thirteen boys ($M = 80.3$ months, $SD = 3.92$) and twelve girls ($M = 82.6$ months, $SD = 4.91$) were tested. An additional 5 children were tested but their data were not used; 3 were due to over-scheduling and 2 were due to inattentiveness during the tasks.

Nonverbal relational judgment task

We used the same nonverbal task used in Experiments 1, 2 and 3. As in Experiment 3, this task was performed second.

Nonverbal Relational Practice Task

We modified the dynamic verbal task used in Experiments 2 and 3 so that children could respond nonverbally. As many parameters as possible were kept the same. For example, children were able to respond immediately and like in the verbal task of Experiment 3, they were not provided any feedback as to the correctness of their responses; however, we replaced the white noise screen (available during verbal response) with the two-colored response screen. The two-colored response screen appeared immediately after the stimulus screen and was available for up to 9s. Like in the previous studies, children were provided with 8 training trials to ensure they understood the instructions, before completing the 48 practice trials.

Results

The critical comparison for this experiment is the performance on the incongruent trials as compared to the previous experiments. We compared the variances as these were shown to be different between Experiments 1 and 2 and Experiment 3. Levene's test for equality of variances for the incongruent trials of Experiment 4 compared to Experiment 3 indicated a significant difference in the variance ($F_{24,24} = 4.3$, $p < .05$), such that children in Experiment 3 performed much more similarly to each other than children in Experiment 4 did. For example, in Experiment 3, 9 children performed perfectly whereas only 4 children performed perfectly in Experiment 4. Similar to children in Experiments 1 and 2, some children performed near 0% correct in Experiment 4 (for at least one relation; $n = 3$) while no child in Experiment 3 performed at 0% correct. A Mann-Whitney U test ($p = .644$) further shows that the distributions are not different between Experiments 1 & 2 and Experiment 4 whereas these were different in comparison to Experiment 3 (see results in Experiment 3). Furthermore, although parametric tests may yield misleading results with the observed bimodal distribution of performance we obtained, we include the test to further

illustrate the difference in means using an independent samples t-test. A comparison between performance on nonverbal incongruent trials of Experiments 1 and 2 vs. Experiment 4 did not yield a reliable difference ($t(df = 78) = -.789$; $p = .433$ (equal variances assumed) mean difference = 5.42; $SE = 7.25$. Cohen's $d = .185$). It seems, then, that practice cannot account for all of the improved performance found in Experiment 3.

One possible criticism of our argument is that our failure to find a significant practice effect in this experiment is because our sample size was too small for the large variability that we observed (even though our sample size in this experiment was the same as it was in Experiment 3), and thus we are basing our argument on a “null” effect. One way to address this issue is to examine effect sizes, such as Cohen's d , which are independent of sample sizes. Cohen's d for the difference between Experiments 1 and 2 and 3 was .513, a “medium” effect. Cohen's d between Experiments 1 and 2 and Experiment 4 is .185, a “small” effect. Another way of visualizing the magnitude of different effect sizes is through a Binomial Effect Size Display (BESD) which is shown in Table 5. This technique, based on Cohen's d or r , reveals the number of participants that would be expected to improve with a “treatment” if 100 were tested (Randolph & Edmonson, 2005). In our case, using labels (in Experiment 3) and practicing without labels (in Experiment 4) can be considered different treatments. These numbers immediately illustrate that seeing the relational stimuli and pressing the button after each trial by itself (i.e., practice without labels) does not lead to as much improvement as practice which includes labeling the stimuli. For example, practice without labels would be expected to lead to improvement in only 8 out of 100 children, while applying labels would be expected to lead to improvement in 24 out of 100 children: a three-fold increase in group performance.

Discussion

In this experiment, children did not apply verbal labels to relations. However, they did see the same number of relational stimuli and responded to each one prior to performing the nonverbal task. However, we did not find an equal amount of improvement on incongruent trials in this experiment. Therefore, we found that practice alone did not improve performance as much as did practice and use of labels. This finding lends support to our conclusion in Experiment 3 that priming children to use verbal labels significantly improved their performance in a subsequent task. Labels, then, seem to provide a mechanism for ignoring salient but inappropriate information, and thus, move to more appropriate (in this case, intrinsic) relational strategies.

General Discussion

The purpose of this study was to shed light on the role of language in the development of the ability to make flexible relational judgments. Towards that end, we examined the relation between performance on verbal and nonverbal judgments of *above*, *below*, *right* and *left* in children from 5 to 10 years of age. Previous work has suggested that language enhances relational reasoning in humans beyond the capacities observed in animals and young children, but there is currently no consensus on exactly how and when language contributes to that enhancement.

Our results from Experiment 1 suggest that acquisition of relational terms does not by itself immediately and automatically improve the ability to make dynamic relational judgments, especially the ones that require ignoring irrelevant information (i.e., overcoming incongruency). Although we found that performance on the verbal task exceeded performance on the nonverbal task in Experiment 1, it did not systematically predict (i.e., correlate with) individual performance on nonverbal incongruent trials. In Experiments 2 and 3 we addressed two possible reasons for these results. In Experiment 2, we asked whether the absence of a predictive correlation between performance on the verbal and nonverbal tasks was because the verbal production task was too easy and did not require the same degree of dynamic and flexible relational coding required by our nonverbal task. Experiment 3 explored the possibility that an unstudied factor was responsible for the increase in nonverbal performance by “dialing up” the accessibility of language. If another factor was responsible, children's performance should not have been affected by the language manipulation.

In Experiment 2 – using a task that more closely paralleled the dynamic nonverbal task of Experiment 1 and included incongruent trials – we found that the ability to overcome incongruency first appeared in the verbal tasks, in which no effects of congruency were found. This finding indicates that using language bolsters the ability to “inhibit” irrelevant information or a dominant response. Perhaps, a greater and greater number of relational contexts are processed through the language lens with development. This idea – that with development more and different relational tasks become unified – was first proposed by Scott et al. (2015). In that work, cluster analyses were used to examine how sixteen different tasks of relational knowledge were performed (similarly or not) by 5- to 10-year-olds, and we found that more tasks were “chunked” together with increasing age. However, it was not clear from that work whether language was the force unifying performance or whether a third factor was responsible.

Why do we believe that the unifying force is language rather than some other (unstudied) processing ability that develops at the same time? In Experiment 3, we dialed up the role of language and found that nonverbal performance improved as a consequence. Performance on the nonverbal task should not have improved if language was not a central force in performance. Furthermore, when we checked for a practice effect in Experiment 4, we found no evidence of one. Finding that increasing language accessibility improved performance in the nonverbal task is strong evidence that language is a causal factor in promoting flexible relational judgments. These findings thus indicate that language is a central force (i.e., a magnet) within the “chunk” that draws in different instantiations of a relation. Our findings also offer an illustration of the concept of “entrenchment” that has been documented in other areas (see e.g., Schmid, 2017). Entrenchment refers to a process of automatization whereby the processing of different elements becomes increasingly holistic (Langacker, 2008). By this view, memory consolidation and chunking are central to increased automaticity. Our findings suggest that as relational terms become more strongly represented, or entrenched, they promote the ability to inhibit irrelevant information across a broader set of contexts.

Our findings are also consistent with the idea that once the last label is learned (e.g., *right* or *left*) then the “set” of possible relations is complete and leads to improvement of

performance on *all* the relations of the set since children who produce *right* and *left* correctly also performed better on the other nonverbal incongruent relations. This idea was first explored in Scott et al. (2015). It could be that fully assembled knowledge of the entire set of labels for the domain (in this case, basic spatial relational terms) provides a solid foundation for making equivalent nonverbal judgments. Like a chair missing a leg, a set that is lacking even one or two labels could make for an unstable, even shifting, platform for making decisions. This could explain why children sometimes switch strategies during a task when they are uncertain of whether they are providing correct answers – which many children did across our experiments and may have been partly responsible for the variability in performance observed. In either case, our findings suggest that language may improve relational judgments beyond what is typically possible by bolstering the ability to ignore salient but irrelevant information. Further corroborating this interpretation is evidence that labels help young children switch their attention in dimensional card-sorting tasks, in which children first sort cards by one dimension such as shape and are then asked to sort by a different dimension such as color (Doebel & Zelazo, 2013).

To summarize, our study extends previous work in four ways. The first is that we examined the ability to make verbal and nonverbal dynamic relational judgments for four relations (*above*, *below*, *right* and *left*) that had not been previously studied together and directly compared (with the exception of our own previous work, Scott et al., 2015). The second way our work differs from past studies is that we examined the ability to make dynamic relational judgments that required intrinsic relational processing. The third is that we examined children's self-generated use of language instead of providing them with particular terms and labels. Finally, children's relational judgments in our tasks were not influenced by non-relational skills (e.g., motor) abilities involved in navigation or re-orientation tasks.

In examining the ability to make all four types of relational judgments together, our study speaks directly to the seemingly inconsistent findings in previous work. By examining verbal and nonverbal abilities in both vertical and horizontal relational planes within the same group of children, our findings may help clarify the reasons behind the mixed results. We found that knowledge of the labels *above* and *below* was firmly in place before knowledge of the labels *left* and *right* and the ability to make nonverbal incongruent relational judgments. Much of the past work suggesting a central role for language has involved the vertical plane (Clark, 1972; Gentner, 2003). Research involving relational judgments along the horizontal plane has yielded much more mixed results. Our findings suggest that knowledge of the words *above* and *below* is clearly established before children can successfully make nonverbal relational judgments across a variety of stimuli, but knowledge of the words *left* and *right* is still developing and appears to be doing so closely in time (slightly earlier but not significantly so) with the nonverbal ability to make incongruent relational judgments. This may help explain why some researchers have found that knowledge of *left* and *right* leads the ability to make relational judgments along the horizontal plane (Hermer & Spelke, 1996; Hermer-Vazquez et al., 2001) whereas other researchers point to other factors (e.g., Ratliff & Newcombe, 2008).

The second way that our work differs from past work is that we were able to examine the ability to make relational judgments across intrinsic and extrinsic frames of reference, which

yielded the same relational response on congruent trials and conflicting ones on incongruent trials. Previous studies that have examined knowledge of these same relations have used tasks that require relational coding with respect to a single, static, extrinsic frame of reference (e.g., Dessalegn & Landau, 2008; Quinn, 2007), much like our congruent trials, which can be solved by a perceptual similarity matching strategy (or mismatching in the case of the dishabituation results from infants). In our incongruent trials, however, children had to be flexible enough to abandon the potentially incorrect but salient extrinsic frame of reference (i.e., the position of the stimulus on the computer screen) for the correct, intrinsic one (i.e., the relation of the dot to the line). In our study, then, if children relied solely on the extrinsic frame of reference (the global position of the stimulus on the screen), then they would solve all of the congruent trials correctly but all the incongruent trials incorrectly. This is essentially a perceptual similarity matching strategy and one that cannot be ruled out in some past work. Only an intrinsic relational strategy, one which involved ignoring extrinsic frames of reference, would lead to perfect performance on both congruent and incongruent trials. We found that both children who had and who had not acquired relational terms could use the intrinsic relational frame but did so in a fragile manner. They used it about 75% of the time for the incongruent trials of Experiments 1 and 2. However, when the use of a verbal code was promoted in Experiment 3, the use of the relational strategy became more robust and less susceptible to interference from the perceptual similarity matching strategy. In fact, correct performance went up to 90% in this experiment. Perhaps children (and nonhuman primates) who perform nearly perfectly on tasks of relational reasoning are spontaneously using symbolic coding to bolster their performance.

The third aspect of our study that is unique is that we allowed children to use any strategy in the nonverbal task, and never directly suggested a specific strategy. We led children to better performance by priming them with a linguistic strategy in a preceding task. In contrast, for example, Loewenstein and Gentner (2005) provided children with a label upfront during their encoding period (which they could use to hold the location in memory) and as the experimenters hid the object on a shelf. Furthermore, by never providing corrective feedback children were free to use a label of their choosing throughout the experimental task, regardless of its preciseness or correctness, and to abandon and switch labels as they found suitable. Thus, our results reflect children's self-generated and spontaneous use of relational terms in nonverbal relational coding rather than an experimenter-guided solution, and is, therefore, more reflective of everyday performance outside of the laboratory. This approach also gave us the opportunity to compare the correctness of self-generated labels to performance on the nonverbal task. One possibility that these findings suggest is that the act of self-generating the tag (any tag) is what improves nonverbal performance.

Our fourth, and final contribution is that our findings complement previous works since our results do not reflect any added demands that might have been factors in previous studies – such as navigation, or object search. Thus, our tasks measure a more isolated ability to make relational judgments. Importantly, the ability to make relational judgments needs to be studied with techniques that capture neural correlates such as fMRI and MEG among others (see Scott et al., 2016, for MEG results from adults on this same task). In studying this paradigm in adults, Scott and colleagues (2016) found that *right-left* judgments manifest themselves differently in neural activity as compared to *above-below* judgments adding

further credence to our developmental findings that *right-left* judgments are cognitively different from *above-below* judgments. Comparatively, the neural evidence from macaques on a nearly identical task involved the prefrontal cortex (PFC; Goodwin et al., 2012), yet adult humans showed no differential neural activation in PFC between the relational planes (Scott et al., 2016). Using our task in conjunction with neural paradigms and with other animals can yield a better picture of how relational judgments develop in humans, how they are represented in the brain, and how they emerge and evolved across different species.

In conclusion, our findings are consistent with the idea that language strengthens basic relational reasoning abilities. Using language to “chunk” spatial relations expands the capacity for relational reasoning, and may free up resources to solve other problems (See Miller, Vlach & Simmering (2016) for similar findings.). Clearly, many questions remain regarding the role of language in the evolution of human cognition, yet these findings offer a solid step towards better understanding the continuities and discontinuities that exist.

Acknowledgments

This project was in part made possible through the University of Minnesota Interdisciplinary Doctoral Fellowship and Doctoral Dissertation Fellowship, as well as an NIH training grant (T32 HD007151) awarded to N.M.S. We thank the families who participated and B. Khalif, A. Hahn, Z. Vang, T. Hermandson-Olson, Y.Y. Cheung, P. Vang, K. Solko, E. Walsh, H. Nauth and T. Xiong for their assistance with conducting the study. We thank A. Georgopoulos for his comments on our manuscript and M. Chafee and D. Boeff for the assistance in the nonverbal task design.

References

- Benedict H. Early lexical development: Comprehension and production. *Journal of Child Language*. 1979; 6:183–200. DOI: 10.1017/S0305000900002245 [PubMed: 468932]
- Brod G, Werkle-Bergner M, Shing L. The influence of prior knowledge on memory: A development cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience*. 2013; 7:139.doi: 10.3389/fnbeh.2013.00139 [PubMed: 24115923]
- Carlson LA, West R, Taylor HA, Herndon RW. Neural correlates of spatial term use. *Journal of Experimental Psychology*. 2002; 28:1391–1408. DOI: 10.1037/0096-1523.28.6.1391 [PubMed: 12542134]
- Casasola M. Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. *Developmental Psychology*. 2005; 41:183–192. DOI: 10.1037/0012-1649.41.1.183. [PubMed: 15656748]
- Christie S, Gentner D. Language helps children succeed on a classic analogy task. *Cognitive Science*. 2013; 38:383–397. DOI: 10.1111/cogs.12099 [PubMed: 24215433]
- Clark EV. On the child's acquisition of antonyms in two semantic fields. *Journal of Verbal Learning and Verbal Behavior*. 1972; 11:750–758. DOI: 10.1016/S0022-5371(72)80009-4
- Clark EV, Hecht BF. Comprehension, production, and language acquisition. *Annual Review of Psychology*. 1983; 34:325–349. DOI: 10.1146/annurev.ps.34.020183.001545
- Cox MV, Richardson JR. How do children describe spatial relationships? *Journal of Child Language*. 1985; 12:611–620. DOI: 10.1017/S0305000900006681 [PubMed: 4077934]
- Dalrymple-Alford EC. Associative facilitation and interference in the Stroop color– word task. *Perception and Psychophysics*. 1972; 11:274–276. DOI: 10.3758/BF03210377
- Dessalegn B, Landau B. More than meets the eye. *Psychological Science*. 2008; 19:189–195. DOI: 10.1111/j.1467-9280.2008.02066.x [PubMed: 18271868]
- Diamond A, Kirkham N, Amso D. Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*. 2002; 38:352–362. DOI: 10.1037/0012-1649.38.3.352 [PubMed: 12005379]

- Doebel S, Zelazo PD. Bottom-up and top-down dynamics in young children's executive function: Labels aid 3-year-olds' performance on the Dimensional Change Card Sort. *Cognitive Development*. 2013; 28:222–232. DOI: 10.1016/j.cogdev.2012.12.001 [PubMed: 24882942]
- Fortes AF, Merchant H, Georgopoulos AP. Comparative and categorical spatial judgments in the monkey: “High” and “low”. *Animal Cognition*. 2004; 7:101–108. DOI: 10.1007/s10071-003-0195-6 [PubMed: 15069609]
- Gava L, Valenza E, Turati C. Newborns' perception of left-right spatial relations. *Child Development*. 2009; 80:1797–1810. DOI: 10.1111/j.1467-8624.2009.01368.x [PubMed: 19930352]
- Gentner D. Why we're so smart. In: Gentner D, Goldin-Meadow S, editors *Language in Mind*. Cambridge, MA: MIT Press; 2003. 195–235.
- Gentner D. Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*. 2010; 34:752–775. DOI: 10.1111/j.1551-6709.2010.01114.x [PubMed: 21564235]
- Gentner D, Anggoro FK, Klibanoff RS. Structure-mapping and relational language support children's learning of relational categories. *Child Development*. 2011; 82:1173–1188. DOI: 10.1111/j.1467-8624.2011.01599.x [PubMed: 21557743]
- Gentner D, Ratterman MJ. Language and the career of similarity. In: Gelman SA, Byrnes JP, editors *Perspectives on Language and Thought: Interrelations in Development*. Cambridge, UK: Cambridge University Press; 1991. 225–277.
- Goodwin SJ, Blackman RK, Sakellaridi S, Chafee MV. Executive control over cognition: Stronger and earlier rule-based modulation of spatial category signals in prefrontal cortex relative to parietal cortex. *The Journal of Neuroscience*. 2012; 32:3499–515. DOI: 10.1523/JNEUROSCI.3585-11.2012 [PubMed: 22399773]
- Halford GS, Wilson WH, Phillips S. Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*. 2010; 14:497–505. DOI: 10.1016/j.tics.2010.08.005 [PubMed: 20884275]
- Haun DBM, Rapold CJ, Janzen G, Levinson SC. Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition*. 2011; 119:70–80. DOI: 10.1016/j.cognition.2012.12.009 [PubMed: 21238953]
- Hayward WG, Tarr MJ. Spatial language and spatial representation. *Cognition*. 1995; 55:39–84. DOI: 10.1016/0010-0277(94)00643-Y [PubMed: 7758270]
- Hermer L, Spelke ES. A geometric process for spatial reorientation in young children. *Nature*. 1994; 370:57–59. DOI: 10.1038/370057a0 [PubMed: 8015605]
- Hermer L, Spelke ES. Modularity and development: The case of spatial reorientation. *Cognition*. 1996; 61:195–232. DOI: 10.1016/S0010-0277(96)00714-7 [PubMed: 8990972]
- Hermer-Vazquez L, Moffet A, Munkholm P. Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition*. 2001; 79:263–99. DOI: 10.1016/S0010-27(00)00120-7 [PubMed: 11165214]
- Hermer-Vasquez L, Spelke ES, Katsnelson AS. Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*. 1999; 39:3–36. DOI: 10.006/cogp.1998.0713 [PubMed: 10433786]
- Hespos SJ, Piccin TB. To generalize or not to generalize: Spatial categories are influenced by physical attributes and language. *Developmental Science*. 2009; 12:88–95. DOI: 10.1111/j.1467-7687.2008.00749.x [PubMed: 19120416]
- Hespos SJ, Spelke ES. Conceptual precursors to language. *Nature*. 2004; 430:453–6. DOI: 10.1038/nature02634 [PubMed: 15269769]
- Klein GS. Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*. 1964; 77:576–588. DOI: 10.2307/1420768 [PubMed: 14255565]
- Kranjec A, Lupyan G, Chatterjee A. Categorical biases in perceiving spatial relations. *PLoS One*. 2014; 9:e98604.doi: 10.1371/journal.pone.0098604 [PubMed: 24870560]
- Landau B, Hoffman JE. Parallels between spatial cognition and spatial language: Evidence from Williams syndrome. *Journal of Memory and Language*. 2005; 53:163–185. DOI: 10.1016/j.jml.2004.05.007
- Landau B, Lakusta L. Spatial representation across species: Geometry, language, and maps. *Current Opinion in Neurobiology*. 2009; 19:12–19. DOI: 10.1016/j.conb.2009.02.001 [PubMed: 19303766]

- Langacker RW. Cognitive grammar: A basic introduction. Oxford University Press; 2008. Ebook version
- Levinson SC. Language and space. *Annual Review of Anthropology*. 1996; 25:353–382. DOI: 10.1146/annurev.anthro.25.1.353
- Levinson SC. Space in Language and Cognition: Explorations in Cognitive Diversity. Cambridge, UK: Cambridge University Press; 2003.
- Li P, Gleitman L. Turning the tables: Language and spatial reasoning. *Cognition*. 2002; 83:265–294. DOI: 10.1016/S0010-0277(2)00009-4 [PubMed: 11934404]
- Loewenstein J, Gentner D. Relational language and the development of relational mapping. *Cognitive Psychology*. 2005; 50:315–353. DOI: 10.1016/j.cogpsych.2004.09.004 [PubMed: 15893523]
- Miller HE, Vlach HA, Simmering VR. Producing spatial words is not enough: Understanding the relation between language and spatial cognition. *Child Development*. 2016; 88:1966–1982. DOI: 10.1111/cdev.12664 [PubMed: 27859021]
- National Research Council. Learning to think spatially. Washington DC: National Academies Press; 2006.
- Pyers JE, Shusterman A, Senghas A, Spelke ES, Emmorey K. 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*. 2010; 107:12116–12120. DOI: 10.1073/pnas.0914044107 [PubMed: 20616088]
- Quinn PC. On the infant's prelinguistic conception of spatial relations. In: Plumert J, Spencer J, editors *The Emerging Spatial Mind*. Oxford: Oxford University Press; 2007. 117–141.
- Randolf JJ, Edmonson RS. Using the Binomial Effect Size Display (BESD) to Present the Magnitude of Effect Sizes to the Evaluation Audience. *Practical Assessment, Research and Evaluation*. 2005; 14:1–7.
- Ratliff KR, Newcombe NS. Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cognitive Psychology*. 2008; 56:142–163. DOI: 10.1016/j.cogpsych.2007.06.002 [PubMed: 17663986]
- Schmid HJ. Entrenchment and the Psychology of Language Learning. Washington, D.C.: American Psychological Association; 2017.
- Scott NM, Leuthold A, Sera MD, Georgopoulos AP. Differential neural activity patterns for spatial relations in humans: A MEG study. *Experimental Brain Research*. 2016; 243:429–441. DOI: 10.1007/s00221-015-4467-6
- Scott NM, Sera MD, Georgopoulos AP. An information theory analysis of spatial decisions in cognitive development. *Frontiers in Neuroscience*. 2015; 9:14.doi: 10.3389/fnins.2015.00014 [PubMed: 25698915]
- Shusterman A, Lee SA, Spelke ES. Cognitive effects of language on human navigation. *Cognition*. 2011; 120:186–201. DOI: 10.1016/j.cognition.2011.04.004 [PubMed: 21665199]
- Smith LB, Rattermann MJ, Sera MD. “Higher” and “lower”: Comparative and categorical interpretations by children. *Cognitive Development*. 1988; 3:341–357. DOI: 10.1016/0885-2014(88)90020-2
- Son JY, Smith LB, Goldstone RL, Leslie M. The Importance of being interpreted: Grounded words and children's relational reasoning. *Frontiers in Psychology*. 2012; 3:1–12. DOI: 10.3389/fpsyg.2012.00045 [PubMed: 22279440]

Highlights

1. We investigated the relation between verbal and nonverbal relational processing.
2. Accessing labels promoted relational coding in a nonverbal spatial task.
3. Early knowledge of the labels did not, by itself, improve nonverbal performance.
4. Any label may do: use of the correct label was not needed to improve performance.
5. Labels enhanced relational coding by helping one ignore irrelevant information.

Figure 1. Panel A depicts the four stimuli used to elicit intrinsic judgments of *above*, *below*, *left* and *right*. Panel B illustrates the four quadrants on the screen where the circle could appear, and that offered extrinsic reference points. Figure originally published in Scott et al. (2015).

Figure 2. Illustration of two different trials in the nonverbal task. The top three panels show an example of an “above” incongruent trial; the bottom three panels show an example of a “left” congruent trial. Touching the red and yellow rectangles would constitute correct responses on these trials. The numbers to the right indicate the amount of time each screen was displayed. Figure as originally published in Scott et al., 2015.

Figure 3.

Illustration of the magnetic tic-tac-toe board used for the production and comprehension verbal tasks in Experiment 1. The house, cat, child and dog illustrate the four positions in which the objects could appear. In this figure, the “cat” is in the “above” position; the “child” is in the “right” position; the “dog” is in the “below” position; and, the “house” is in the “left” position. This figure was originally published in Scott et al. (2015).

Figure 4. Children's performance on the verbal production (left panel) and comprehension task (right panel). Only the comprehension task could be measured against chance, which was set at 37.5% (for 3 out of 8 possible correct locations) and is illustrated by the dotted line.

Figure 5.

Average percent correct on the nonverbal tasks of Experiment 1 for each age group. Congruent and incongruent trials of the nonverbal task are displayed separately. Error bars illustrate a standard error of ± 1 of the means. Chance performance was 50% and is illustrated by the black line.

Figure 6. Scatterplots of individual children's performance on the nonverbal incongruent trials (filled circles) and verbal production task (unfilled triangle) as a function of age in months. Each relation is represented in a separate plot. The trajectory of the lines is determined by the performance of 20% of the children and is calculated by a moving window. The curve for development of verbal production is represented by the solid line, and development on nonverbal incongruent trials is represented by the dotted line.

Figure 7. Illustration of the computer screens presented in the dynamic verbal task used in Experiments 2 and 3. Two trials are depicted: first is an above-congruent trial and the second is a right-incongruent trial. The duration of each screen is written to the right of the corresponding screen.

Figure 8. Error plot of percent correct on the verbal task in Experiment 2. Congruent and incongruent trials are displayed separately. Error bars illustrate ± 1 standard error of the means.

Figure 9. Box plot of percent correct on nonverbal and verbal tasks for 6-7.5-year-olds in Experiments 1 and 2. Congruent and incongruent trials of the nonverbal task are displayed separately, but congruent and incongruent trials of the verbal task are averaged for an overall score for Experiment 2. The median is displayed as a black line inside the box. Outliers are displayed as a star symbol and a numerical tag.

Figure 10.

Error plot of performance on nonverbal and verbal tasks of Experiment 3. Congruent and incongruent trials are displayed separately. Error bars illustrate ± 1 standard error of the means.

Figure 11.

Boxplots of performance by children on the nonverbal incongruent trials across Experiments 1 and 2 and Experiment 3. The variability in performance was significantly different between the three experiments. The median is displayed as a black line inside the box. Outliers are displayed as a star symbol and a numerical tag (to include overlapping scores).

Table 1

Descriptive information about the participants

Age (year)	Girls			Boys		
	<i>M (months)</i>	<i>Range</i>	<i>SD</i>	<i>M (months)</i>	<i>Range</i>	<i>SD</i>
5	66.3	60-71	4.03	65.2	61-70	2.74
6	77.2	72-83	3.33	77.2	72-83	3.49
7	90.6	86-95	3.92	89.3	84-95	3.62
8	101.3	97-106	3.06	101.2	96-106	4.32
9	113.2	109-118	3.01	113.7	109-117	2.79
10	126	122-131	3.30	124.5	120-128	2.59

Note. Each group was comprised of 10 girls and 10 boys for a total of 20 children per age group.

Table 2
Spearman's Rho correlation between percent correct on verbal and nonverbal incongruent trials

Nonverbal Incongruent trials	Age	Verbal Production			
		Above	Below	Right	Left
Above	5	.26	.26	-.16	-.10
	6-7	-.13	-.13	.17	.28
	8-10	--	--	.30*	.17
Below	5	.03	.03	.12	.18
	6-7	-.13	-.13	.20	.37*
	8-10	--	--	-.11	-.11
Right	5	.18	.18	-.14	-.09
	6-7	-.17	-.17	.13	.22
	8-10	--	--	-.11	-.11
Left	5	.39	.39	.01	.06
	6-7	-.14	-.14	.17	.27
	8-10	--	--	.09	.26*

Note. Squares indicate where significant correlations are expected if acquisition of specific terms is driving nonverbal performance; the dashed lines indicate relations in which no correlations could be computed due to lack of variability in one variable (e.g., perfect production of above or below).

* Correlation is significant at the 0.05 level (2-tailed)

[^] Sample sizes for the ages are: 5-year-olds N = 20, 6-7-year-olds N = 40, 8-10-year-olds N = 60

Table 3
Spearman's Rho correlations between percentage correct on nonverbal incongruent trials and verbal production congruent and incongruent trials in Experiment 2

Nonverbal-Incongruent	Production-Congruent				Production-Incongruent			
	Above	Below	Right	Left	Above	Below	Right	Left
Above	.16	.50*	-.02	-.06	-.14	--	-.12	-.00
Below	.12	.26	.09	.09	-.15	--	-.03	-.16
Right	.06	.41*	.19	.14	-.19	--	.05	.23
Left	.31	.36	.03	-.05	-.17	--	-.08	.10

Notes. Squares indicate where significant correlations are expected if language acquisition is driving performance. .

The dashed lines indicate relations in which no correlations could be computed due to lack of variability in one variable (e.g., perfect production of below).

* Correlation is significant at the 0.05 level (2-tailed)

Table 4
Spearman's Rho correlation between percentage correct performance on nonverbal incongruent trials, and verbal production congruent and incongruent trials in Experiment 3

Nonverbal-Incongruent	Production-Congruent				Production-Incongruent			
	Above	Below	Right	Left	Above	Below	Right	Left
Above	.08	-.13	.16	.14	-.14	-.28	.13	.10
Below	-.05	-.25	.56*	.54*	-.34	-.34	.62*	.46*
Right	.17	-.14	.39	.42*	-.11	-.27	.44*	.34
Left	.33	.25	.24	.30	.27	.10	.33	.25

Note. Squares indicate where significant correlations are expected if knowledge of the precise relational term is driving performance.

* Correlation is significant at the 0.05 level (2-tailed)

Table 5
Binomial Effect Size Displays for the use of labels in Experiment 3 and practicing in Experiment 4. The numbers represent the number of participants out of 100 expected to improve by using labels and practicing

Labeled Relations (Expt. 3)		
$d = .513$; $r = .2341$	Improved	Not Improved
Used Labels (+practiced)	62	38
Did Nothing	38	62
Practiced without labels (Expt. 4)		
$d = .185$; $r = .0865$	Improved	Not Improved
Practiced without labels	54	46
Did Nothing	46	54