

Spatial Metaphor and the Development of Cross-Domain Mappings in Early Childhood

Ariel Starr * & Mahesh Srinivasan

In press at *Developmental Psychology*

Department of Psychology, University of California, Berkeley

*** Correspondence:** Ariel Starr, Department of Psychology, UC Berkeley, 134 Barker Hall, Berkeley, CA 94720, arielstarr@berkeley.edu

Abstract

Spatial language is often used metaphorically to describe other domains, including time (*long* sound) and pitch (*high* sound). How does experience with these metaphors shape the ability to associate space with other domains? Here, we tested 3- to 6-year-old English-speaking children and adults with a cross-domain matching task. We probed cross-domain relations that are expressed in English metaphors for time and pitch (length-time and height-pitch), as well as relations that are unconventional in English but expressed in other languages (size-time and thickness-pitch). Participants were tested with a *perceptual matching task*, in which they matched between spatial stimuli and sounds of different durations or pitches, and a *linguistic matching task*, in which they matched between a label denoting a spatial attribute, duration, or pitch, and a picture or sound representing another dimension. Contrary to previous claims that experience with linguistic metaphors is necessary for children to make cross-domain mappings, children performed above chance for both familiar and unfamiliar relations in both tasks, as did adults. Children's performance was also better when a label was provided for one of the dimensions, but only when making length-time, size-time, and height-pitch mappings (not thickness-pitch mappings). These findings suggest that, although experience with metaphorical language is not necessary to make cross-domain mappings, labels can promote these mappings, both when they have familiar metaphorical uses (e.g., English 'long' denotes both length and duration), and when they describe dimensions that share a common ordinal reference frame (e.g., size and duration, but not thickness and pitch).

Keywords: metaphor; linguistic relativity; cross-modal matching; cognitive development

Spatial Metaphor and the Development of Cross-Domain Mappings in Early Childhood

Across languages and cultures, spatial language is often used metaphorically to describe other domains (Alverson, 1994; Lakoff & Johnson, 1980; Sweetser, 1991). In English, for example, we describe temporal durations as *long* or *short*, numbers as *big* or *small*, and auditory pitches as *high* or *low*. It is widely agreed that these spatial metaphors are not just facts about language, but also reflect non-linguistic mappings between representations of space and of other domains, such as duration and pitch (e.g., Bueti & Walsh, 2009; Cantlon, Platt, & Brannon, 2009; Lourenco & Longo, 2011; Walsh, 2003). Further, by some accounts, such mappings play an important role in human cognition, and may allow us to learn and reason about abstract domains, like time, using more concrete domains, like space, for which we have more direct sensory experience (Boroditsky, 2000; Casasanto, 2010; Gibbs, 1994; Lakoff & Johnson, 1980). Given their theoretical importance, it is critical to characterize the factors that shape how cross-domain mappings arise in development. Here, we build on prior studies (e.g., Casasanto, Fotakopoulou, & Boroditsky, 2010; de Hevia, Vanderslice, & Spelke, 2012; Dolscheid, Hunnius, & Majid, 2015; Shayan, Ozturk, Bowerman, & Majid, 2014; Smith & Sera, 1992) to explore the role of linguistic experience in the development of space-time and space-pitch mappings in three- to six-year-old English-speaking children and adults.

Although spatial metaphors for time and pitch are common across languages, there is also variety in the exact spatial relations invoked. Prior studies have exploited this cross-linguistic variation to test hypotheses about the role of linguistic experience in the development of cross-domain associations (Boroditsky, 2000; 2001; Dolscheid et al., 2015; Dolscheid, Shayan, Majid, & Casasanto, 2013; Shayan et al., 2014). For example, in English, temporal durations are described in terms of two-dimensional length, whereas languages including Greek and Spanish

use three-dimensional spatial terms (i.e., such that ‘larger’ sounds are longer in duration). Likewise, though English describes pitch in terms of height, languages including Turkish and Farsi use terms related to thickness (i.e., such that ‘thicker’ sounds are lower in pitch). Notably, evidence from adults suggests that this cross-linguistic variation predicts variation in behavioral responses from adult speakers of different languages (e.g., Boroditsky, 2000; 2001; Casasanto, 2008; Dolscheid et al., 2013; Dolscheid & Casasanto, 2015). In particular, although adults can form cross-domain mappings in explicit tasks across a range of different domains (Marks, 1978; Shayan et al., 2014; Stevens, 1957), studies employing more implicit measures suggest that familiarity with language-specific metaphors may affect the automaticity with which cross-domain associations are accessed (Casasanto, 2008; Dolscheid et al., 2013).

For example, in one study, adult speakers of Dutch, which uses height-pitch metaphors, had to vocally reproduce the tone of a pitch that was presented simultaneously with a line that varied either in height or in thickness. Interestingly, Dutch-speaking subjects’ reproductions of pitch were only biased by irrelevant height information and not by irrelevant thickness information (Dolscheid et al., 2013). The reverse pattern of results – a biasing effect of thickness but not height on pitch reproduction – occurred for Farsi speakers (Dolscheid et al., 2013). These findings suggest that Dutch-speaking adults automatically process a height-pitch mapping, while Farsi-speaking adults automatically process a thickness-pitch mapping, consistent with the different spatial metaphors expressed in their respective languages. Other studies have reported similar findings for the case of space and duration. In particular, while adult speakers of English, which uses length-duration metaphors, are biased by irrelevant variation in spatial length when estimating duration (but not by irrelevant volumetric information), adult speakers of Greek, which uses size-duration metaphors, are biased by irrelevant volumetric information (but not by

variation in spatial length (Casasanto, 2008). Consistent with the idea that experience with linguistic metaphors causes the cross-domain interference documented in these previous studies, brief training in using novel metaphors – e.g., in which Dutch speakers practice using the language of thickness to describe pitch (Dolscheid et al., 2013) – results in subsequent cross-domain interference, e.g., such that Dutch speakers are affected by irrelevant thickness information when reproducing pitch (see also Casasanto, 2008).

Although the findings reviewed above suggest that experience with linguistic metaphors may shape cross-domain mappings, this does not entail that linguistic experience is required for constructing these mappings, e.g., such that children only learn to associate thickness with pitch by learning that a word for thickness also labels pitch. Indeed, a large body of work has shown that, in looking-time studies, infants are sensitive to multiple types of cross-domain correspondences, including mappings that are not encoded by metaphors in infants' linguistic environments (de Hevia & Spelke, 2010; de Hevia, Izard, Coubart, Spelke, & Streri, 2014; Dolscheid, Hunnius, Casasanto, & Majid, 2014; Fernández-Prieto, Navarra, & Pons, 2015; Lewkowicz & Minar, 2014; Lourenco & Longo, 2010; Mondloch & Maurer, 2004; Srinivasan & Carey, 2010; Walker et al., 2010). For example, in one study, 4-month-old infants growing up in a Dutch-speaking environment – which includes height-pitch but not thickness-pitch metaphors – preferred to look at animations that were increasing in visuospatial height or decreasing in thickness when listening to a sliding tone that increased in pitch, and preferred to look at animations that were decreasing in visuospatial height or increasing in thickness when listening to a sliding tone that decreased in pitch (Dolscheid et al., 2014). Similarly, other studies have suggested that infants associate stimuli of longer spatial length or greater size with stimuli of longer temporal duration (Lourenco & Longo, 2010; Srinivasan & Carey, 2010).

Together, findings from the infant literature suggest that experience with the metaphors of one's language is not necessary for forming cross-domain mappings. Instead, infants may initially have access to a wide range of cross-domain mappings – which could be innate or constructed from early experience – and experience with language-specific metaphors over development may selectively strengthen some mappings but not others, such that by adulthood, speakers of a language automatically access only those cross-domain mappings that are reinforced by their language (Casasanto, 2008; 2010; Hespos & Spelke, 2004). Critically, this account would predict that with increasing age and exposure to language-specific metaphors, children should find it easier to make cross-domain mappings that reflect language-specific metaphors and more difficult to make mappings that are absent from their language. Thus, one would expect a developmental trajectory that proceeds from the infant starting point of sensitivity to many cross-domain associations to the adult end point of sensitivity – at least on implicit tasks – to only those cross-domain mappings that are reinforced by language.

Although many studies have probed infants' and adults' sensitivity to cross-domain mappings, comparatively few have explored pre-school aged children's ability to make cross-domain mappings (de Hevia et al., 2012; Dolscheid et al., 2015; Shayan et al., 2014; Smith & Sera, 1992), which limits conclusions about the effect of experience with linguistic metaphors on cross-domain mappings. One exception is a study by Shayan and colleagues (2014), which investigated the development of thickness-pitch mappings in two- to four-year-old children who spoke German, which uses height metaphors for pitch, or either Farsi or Turkish, both of which use thickness metaphors for pitch. We describe their method in detail below, which we adapted for the present study.

To begin, in what the authors called the *linguistic similarity matching task*, children first heard either a verbal description of a pitch (a snake's voice was described as 'high' or 'low' in German, or 'thick' or 'thin' in Farsi and Turkish), or of the thickness of a snake (it was described as 'thick' or 'thin' using the relevant words from the three languages), without any presentation of an actual sound or snake. Children were then asked to map the referent of the description to one of two presented stimuli from the other dimension that were not verbally described. Thus, in some trials, children were told that a 'thick' (or 'thin') snake was looking for its friend that looked much like it, and then had to guess which of two houses the friend was in by listening to the sounds that came from each house. The children were then played the two sounds that varied in pitch, and these sounds were not verbally described. In other trials, if children had initially heard a description of the pitch of the snake's voice, they were told that the snake was calling out to a friend that was much like it, and were then shown two snakes that varied in thickness (but their thickness was not verbally described) and asked which was the snake who was being called out to.

The authors reasoned that because Farsi and Turkish children speak a language in which the same words are used for thickness and pitch, the verbal descriptions might give them an advantage – relative to German-speaking children – by cueing them into how variation in the attributes of the presented stimuli were relevant to how to perform the mappings, i.e., allowing Farsi-speaking children to reason that a 'thick' snake probably has a friend that makes a sound that children would also describe as 'thick', as opposed to a 'thin' sound. To test whether labels were necessary for children to make consistent mappings, the authors also included a second, *non-linguistic similarity matching task*, which was similar except that no linguistic cues were provided. For example, in the thickness-to-pitch trials, participants were shown a snake that was

either thick or thin (but was not described as such) and were told that it was looking for its friend, and had to decide which house its friend was in by listening to the two sounds. In the pitch-to-thickness trials, children were told that a snake was calling out to its friend and were played the sound. They were then shown two snakes that varied in thickness (but were not described as such) and chose which was the snake that was being called on.

Strikingly, Turkish- and Farsi-speaking children reliably mapped thick objects to low pitches and thin objects to high pitches, whereas German-speaking children demonstrated no consistent mapping. This was true not just in the linguistic matching task, but also in the non-linguistic matching task, in which no verbal cues were provided. These findings are consistent with the idea that experience with thickness-pitch linguistic metaphors strengthens pitch-thickness mappings in Farsi- and Turkish-speaking children, who are exposed to these metaphors, but not in German-speaking children, who are exposed to height-pitch metaphors. However, it remains possible that other cultural or environmental differences between the German, Farsi and Turkish samples – beyond differences in children’s exposure to thickness-pitch linguistic metaphors – were responsible for the different responses of these children. Indeed, another study using different stimuli found that a sample of five-year-old Dutch-speaking children failed to make consistent height-pitch mappings in a nonlinguistic matching task, despite the presence of height-pitch metaphors in the Dutch language (Dolscheid et al., 2014). These conflicting findings highlight the importance of testing children’s cross-domain mappings for both familiar and unfamiliar metaphors within the same population and using the same tasks, which was a goal of the present study.

Finally, although the findings of Shayan and colleagues (2014) open the possibility that children’s ability to map between thickness and pitch is shaped by their exposure – or lack

thereof – to a thickness-pitch metaphor in their linguistic environment, it is unclear whether this possible effect of experience with language-specific metaphors might extend to other cross-domain associations. Mappings between space and pitch may be especially likely to require scaffolding from language because they do not share a common representational format. Specifically, unlike dimensions such as number, length, size, and duration, which can all be represented and described in terms of *more* versus *less*, differences in pitch are qualitative and vary in frequency rather than in amount (i.e., pitch is a *metathetic* dimension, Stevens, 1957). Thus, while mappings between space and time could be facilitated by their common reference frame, whereby *more* in space can be mapped to *more* in time, mappings between space and pitch cannot, because it is unclear how to interpret *more* in the context of pitch (Mondloch & Maurer, 2004; Smith & Sera, 1992). Because space-pitch mappings may rely more strongly on experience with specific linguistic metaphors than space-time mappings, it is important to test both types of mappings, which was a second goal of our study.

The Present Studies

Here, we expand on previous studies of the development of cross-domain associations by testing both space-time and space-pitch associations, and by directly comparing associations corresponding to familiar versus novel metaphors in the same population. Additionally, we test a large number of English-speaking children between the ages of three and six to gain insight into the fine-grained developmental trajectory of cross-domain associations. Our study adapts the general cross-domain matching tasks used by Shayan and colleagues (2014): We present participants with a game in which they need to match between pictures of aliens and the sounds they produce, while varying the perceptual attributes of the aliens and sounds, to focus on the cross-dimensional relations of interest (i.e., length-duration, size-duration, height-pitch, and

thickness-pitch). As in these previous studies, we use both a perceptual matching task, in which participants match between the aliens and sounds without any linguistic cues, and a linguistic matching task in which one dimension is verbally described and matched to the aliens or sounds. Our study design focuses on three critical factors that could contribute to the development of children's cross-domain associations.

First, we investigate the effect of experience with specific spatial metaphors by testing whether English-speaking children are better able to match across domains in ways that reflect familiar spatial metaphors compared to unfamiliar metaphors. For time, we compared children's ability to map between length and duration (familiar relation) and between size and duration (unfamiliar relation). For pitch, we compared children's ability to map between height and pitch (familiar relation) and between thickness and pitch (unfamiliar relation). If experience with specific metaphors selectively strengthens some cross-domain mappings but not others, children should begin to show an advantage for mappings that correspond to familiar metaphors as they get older and accrue more experience with language-specific metaphors (Casasanto, 2008; 2010). This familiarity advantage could be present regardless of whether children are matching between verbal descriptions and pictures or sounds (i.e., in the linguistic matching task) or between unlabeled pictures and sounds (i.e., in the perceptual matching tasks).

Second, we compare cross-domain associations between space and time to associations between space and pitch. We focus on associations with time and pitch because while both time and pitch are frequently described using spatial language, they have different dimensional structures, as previously noted. In particular, the shared structure of space and time may provide a specific advantage for cross-domain space-time mappings that does not extend to space-pitch mappings. The effect of metaphor familiarity may also interact with whether children are

mapping space to time or space to pitch. By this account, children could make space-time mappings due to their shared structure, regardless of whether they are encoded by English metaphors. In particular, children could recognize that, just as something that is big is more than something that is small, something that is long in time is more than something that is short, allowing them to map a big object onto a long temporal duration even though this metaphor doesn't exist in English. In contrast, because space and pitch do not share a common representational format, experience hearing linguistic metaphors – e.g., that 'high' and 'low' refer to both spatial height and pitch – may be more important for these mappings.

Finally, we were interested in whether children might be better at matching across domains in the linguistic matching task, when linguistic descriptions are provided for one of the domains, compared to in the perceptual matching task, when no verbal cues are provided. As noted above, the presence of a label could potentially cue children into how to perform a cross-domain mapping, especially when the label has familiar, metaphorical uses. For example, if children are asked to match an alien described as 'long' to one of two sounds that vary in duration, they could think that the sound that they might also describe as 'long' in duration would be the best choice.

Further, the advantage provided by labels could extend not just to familiar metaphorical words in English, like 'long', but also to spatial words like 'big', which do not typically apply to duration in English. Previous research suggests that children can form some cross-domain perceptual mappings by initially forming mappings between labels from different domains according to their ordinal relations (Smith & Sera, 1992). For example, if a child hears an alien described as 'big' in size, and are presented with two sounds that they would describe as 'long' and 'short' in duration, they could map the word 'big' to the word 'long', because just as 'long'

is more than ‘short’, ‘big’ is more than ‘small’. Relatedly, children could link ‘big’ to ‘long’ because both of these adjectives are *unmarked* (e.g., one typically asks ‘how big’ or ‘how long’ something is, whereas asking ‘how small’ or ‘how short’ something is presupposes its smallness or shortness; e.g., Clark, 1973). These cues could help the child infer that the ‘big’ alien likely makes the long sound. Interestingly, however, these strategies might lead children astray when mapping between an alien described as ‘thick’ and sounds the child would describe as ‘high’ or ‘low’ in pitch. Because ‘thick’ is more than ‘thin’, and ‘high’ is more than ‘low’ – and because ‘thick’ and ‘high’ are unmarked adjectives – children might guess that a ‘thick’ alien should be linked with a ‘high’ sound and a ‘thin’ alien with a ‘low’ sound, which is the opposite of how thickness-pitch mappings are encoded in Farsi and Turkish metaphors.

Methods

Participants

The participants of this study were 128 children aged 3 to 6 years (3-year-olds: N = 32, mean age: 3.48 years, range: 3.02-3.98 years, 21 female; 4-year-olds: N = 32, mean age: 4.54 years, range: 4.02-4.94 years, 10 female; 5-year-olds: N = 32, mean age: 5.34 years, range: 5.02-5.95 years, 17 female; 6-year-olds: N = 32, mean age: 6.58 years, range: 6.00-6.98 years, 17 female) and 16 adults (mean age: 21.25 years, range: 18.63-27.59 years, 12 female). All participants were native English language speakers from the Berkeley, California area who were not regularly exposed to or fluent in a second language. Child participants were tested in the lab or in a quiet area of a preschool or children’s museum. All adult participants were tested in the lab. Data from an additional 14 children and one adult were excluded from analyses due to failure to complete the experiment (nine children), inattention (two children), experimenter error (three children), or performance more than 3 standard deviations below the group mean (one

adult). The demographics of the participants were representative of the Berkeley area, which is approximately 50% White, 20% Asian, 10% Black or African American, 10% Hispanic or Latino, and 6% from two or more races. Adult participants and guardians of child participants gave written consent to protocol #2013-08-5546 “Language and Cognition in Children and Adults,” which was approved by the UC Berkeley Committee for Protection of Human Subjects. Adult participants were compensated with university course credit and children received a small gift.

Materials

Spatial stimuli consisted of pictures of cartoon aliens that differed in length, overall size, vertical position, or thickness (Figure 1). Temporal and pitch stimuli consisted of monotonic tones that differed in duration or auditory pitch, respectively. Tones that differed in duration had a constant pitch of 384 Hz and were either 1 second or 3 seconds in length. Tones that differed in pitch had a constant duration of 2 seconds and a pitch of either 256 Hz or 512 Hz. Visual stimuli were created using Adobe Photoshop, and auditory stimuli were created in Audacity using sine waveforms. All stimuli were presented using a Macintosh laptop computer.

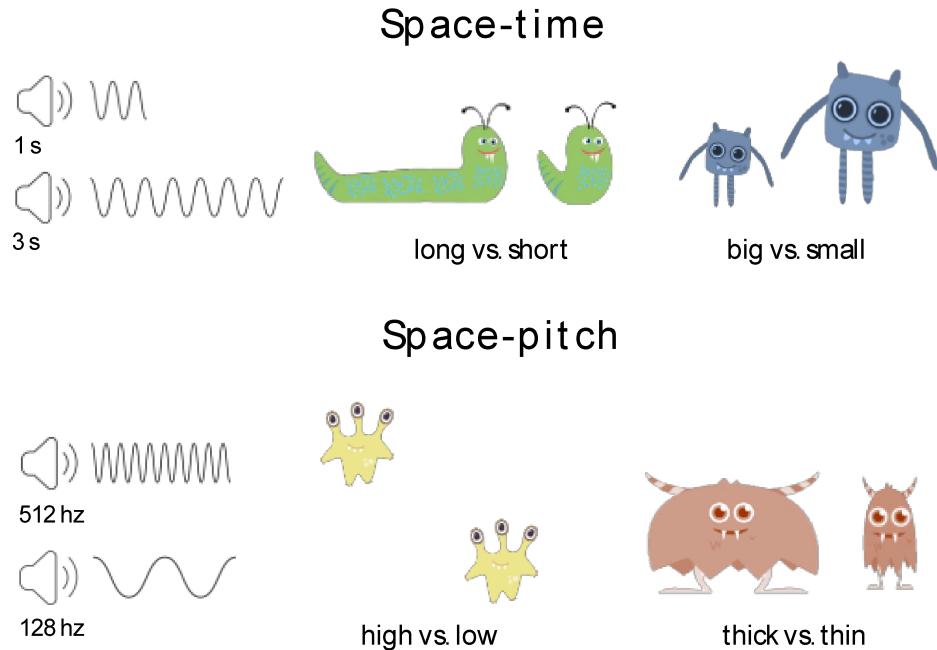


Figure 1. Stimuli used in all conditions.

Procedure

All experimental sessions began with four familiarization trials (described below), before proceeding to the test blocks. Each test block probed participants' ability to perform a cross-dimensional mapping, using both the *perceptual matching task* and *linguistic matching task*. The four test blocks consisted of two space-time blocks and two space-pitch blocks, and tested both familiar relations in English (length-duration for space-time and height-pitch for space-pitch), and unfamiliar relations (size-duration for space-time and thickness-pitch for space-pitch). Within each test block, the perceptual matching task always preceded the linguistic matching task to minimize the likelihood that participants would spontaneously use labels in the perceptual matching task (see *Supplementary Material* for scripts).

Adult participants completed all four test blocks. We counterbalanced whether participants received the two space-time blocks first, or the two space-pitch block first. Within the space-time and space-pitch blocks, the unfamiliar relation (e.g., size-duration or thickness-

pitch) always preceded the familiar relation (e.g., length-duration or height-pitch) to ensure that immediate prior experience with a familiar mapping would not bias adults' interpretation of the unfamiliar mappings. In contrast to adults, child participants completed only two blocks overall: One block testing space-time mappings and one block testing space-pitch mappings. For each block (space-time or space-pitch), children were randomly assigned to receive either the familiar relation or the unfamiliar relation. The order of the two blocks was counterbalanced across child participants.

Familiarization trials. Following Shayan et al. (2014), the familiarization trials were designed to have the same structure as the test trials, and involved matching pictures of animals to the sounds they make. In the first two trials, two animals were displayed on the screen (e.g., a cow and a bird). An animal sound was then played (e.g., a cow mooing), and the participant was instructed to point to the animal that makes that sound. In the second two trials, one animal appeared centrally on the screen in front of two trees (e.g., a cow). The participant was told that the animal was looking for another animal just like it that was hiding behind one of the trees, and played a sound from each tree (e.g., the sounds of a cow and a bird) while the tree wiggled. The participant was instructed to point to the tree where the animal that looks like the visible animal was hiding. Feedback was given if necessary, and all participants answered all four questions correctly before moving onto the test trials.

Test trials. Because each test block probed participants' ability to map between stimuli that varied in one dimension (e.g., a long vs. short alien) to stimuli that varied in another dimension (e.g., a long vs. short tone), we first familiarized participants to the stimuli that would be presented in that block. To accomplish this, at the beginning of each test block, participants were first shown pictures of both relevant aliens and listened to both types of sounds.

Participants were prompted to notice the difference between the exemplars (e.g., ‘See these two aliens? Can you see that they’re different?’), but specific attributes were not highlighted or labeled by the experimenter.

Perceptual matching task. In the perceptual matching task, participants matched between pictures of aliens and the sounds they make (see Figure 2, top row). Critically, in these trials labels were never used to describe the stimuli (see *Supplementary Material*). There were two trial types: 1) *space as source* trials in which participants were shown an alien that had a particular spatial attribute (e.g., it was long or short, big or small, high or low, or thick or thin) and had to guess which of two sounds that varied in duration (for length and size) or pitch (for height and thickness) was made by an alien that ‘looks just like’ the first alien, 2) *space as target* trials in which participants were presented a sound that had a particular duration (e.g., it was long or short) or pitch (e.g., it was high or low) and had to guess which of two aliens that varied in length or size (for duration), or height or thickness (for pitch) ‘looks like’ the alien that had made the sound.

Within each test block (e.g., length-duration, size-duration, etc.), there were two types of *space as source* trials, which varied in which pole of the spatial attribute was exemplified by the first, presented alien (e.g., whether the alien was long or short), and two *space as target* trials, which varied in which pole of the temporal or pitch attribute was exemplified by the first presented sound (e.g., whether the tone was long or short). For each test block, child participants were tested on each of the four types of *space as source* and *space as target* trials once, while adult participants were tested on each of these trial types twice. We counterbalanced the order of the *space as source* and *space as target* trials, which pole of the attribute was tested first (e.g., long or short), and the left/right position of the aliens and sounds.

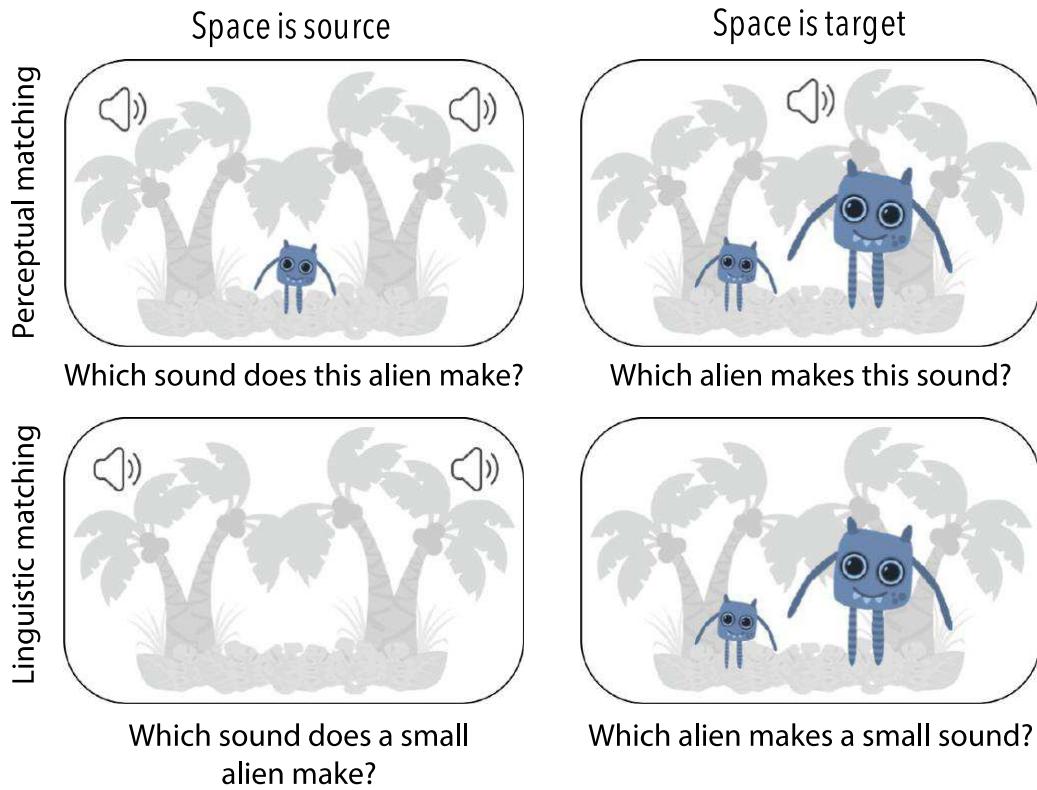


Figure 2. Example trials for the size-time condition. Perceptual matching trials are illustrated the top row, linguistic matching trials are illustrated in the bottom row.

Linguistic matching task. The linguistic matching task was similar to the perceptual matching task except that, instead of presenting the first alien or sound, a label was used to describe it. The participant was asked to indicate which of two presented sounds is made by the described alien, or which of two presented aliens makes the described sound (see Figure 2, bottom row; see *Supplementary Material* for scripts). In the *space as source* trials, the spatial dimension was described and participants chose one of two auditory matches. For example, in the length-duration condition, the experimenter said, “A *long* (or *short*) alien is looking for another alien just like him”, and then played two sounds and asked the participant “Which sound

does a *long* (or *short*) alien make?" In the *space as target* trials, the type of sound was described and participants chose which of two visually presented aliens makes that type of sound. The experimenter said, "Here are two more aliens! Which one do you think makes a *long* (or *short*) sound?" The aliens and sounds were described as 'long' or 'short' in the length-duration trials, 'big' or 'small' in the size-duration trials, 'high' or 'low' in the height-pitch trials, and 'thick' or 'thin' in the thickness-pitch trials.

As in the perceptual matching task, within each test block, child participants received two *space as source* trials, which varied in which pole of the spatial attribute was described (e.g., whether the alien was described as 'long' vs. 'short'), and two *space as target* trials, which varied in which pole of the temporal or pitch attribute was described (e.g., whether the tone was described as 'long' or 'short'). Adult participants received two trials of each of the four types of *space as source* and *space as target* trials. We counterbalanced the order of the *space as source* and *space as target* trials, which pole of the attribute was tested first (e.g., long or short), and the left/right position of the aliens and sounds. For both the perceptual matching and linguistic matching tasks, our analyses collapsed across responses from *space is source* and *space is target* trials, as well as across trials that tested different poles of the dimensions (e.g., long and short).

Sample size justification

Power analyses were conducted to determine the appropriate sample size based on the size of the main effect of language in Shayan et al. (2014), which maps most closely to our primary aim of assessing the effect of metaphor familiarity. The effect sizes for language in this previous study were medium to large (ds between .6 and 1). To detect a similarly sized effect of familiarity in the present study with $d = .6$, $\alpha = .05$, and power $(1-\beta) = .8$, 45 participants are needed for each level of familiarity. Collapsing across age groups, our sample includes 64

children in the familiar and unfamiliar conditions for both the space-time and space-pitch dimensions, which indicates that our study is sufficiently powered to detect an effect of familiarity. Our larger sample size provides us with the statistical power necessary to detect potentially smaller effects of dimension and age, factors for which previous estimates of effect size are not available.

Results

Participants' choices for both the perceptual matching and linguistic matching tasks were coded as 'correct' if their choices reflected matching according to familiar spatial metaphorical relations that exist in English (i.e., matching a long or short alien to a long or short tone, respectively) or unfamiliar metaphorical relations that exist in other languages (i.e., matching a big or small alien to a long or short tone, respectively). Beginning with adults, matching performance was near ceiling across the two tasks (96.2%). These results suggest that English-speaking adults are readily able to perform cross-dimension space-time and space-pitch mappings, regardless of whether these mappings are encoded in English or whether a label is provided for one of the dimensions in the task, which is consistent with previous findings that adults can perform a variety of cross-dimension mappings in explicit tasks (Marks, 1978; Shayan et al., 2014; Stevens, 1957).

A primary question of our study was whether children would make consistent cross-domain mappings, even for relations that are not reflected in English metaphors. Strikingly, children overall matched at above chance levels (50%) for all cross-domain mappings (length-time = 73.0% [SE = 2.34%], size-time = 70.1% [SE = 2.40%], height-pitch = 67.% [SE = 2.21 %], thickness-pitch = 59.6% [SE = 2.49%]; one-sample t-tests comparing performance to chance,

all $ts > 3.55$, $ps < .001$). Critically, these results are unexpected on the hypothesis that experience with linguistic metaphors is necessary for children to make cross-domain mappings (Shayan et al., 2014), because English-speaking children demonstrated consistent mappings between size and duration, and between thickness and pitch, which are not reflected in English metaphors.

To better understand children's performance, we conducted a repeated measures ANOVA with match type (perceptual or linguistic) as a within-subject factor and dimension (space-time or space-pitch), familiarity (familiar or unfamiliar), and age group (three, four, five, or six years) as between-subjects factors. The model yielded a significant main effect of dimension ($F(1, 240) = 7.45$, $p < .01$, $\eta^2 = .02$), which reflected that children overall performed more accurately for space-time mappings compared to space-pitch mappings (space-time mean accuracy: 78.5% [SE = 1.68%]; space-pitch: 68.9% [SE = 1.67%]). There was also a significant effect of match type ($F(1, 240) = 39.8$, $p < .001$, $\eta^2 = .05$), which reflected that children were overall better able to perform cross-domain mappings in the linguistic matching task (mean accuracy: 73.7% [SE = 1.62%]), when one of the domains was described by a label, compared to in the perceptual matching task, where no labels were provided (mean accuracy: 61.3% [SE = 1.70%]). We also observed a clear developmental trajectory in children's performance, as reflected by a significant main effect of age ($F(3, 240) = 9.50$, $p < .001$, $\eta^2 = .08$). Older children performed better than younger children, with the largest jump in performance occurring between ages three and four (mean accuracy in 3-year-olds: 54.5% [SE = 2.29%]; 4-year-olds: 68.2% [SE = 2.24%]; 5-year-olds: 72.3% [SE = 2.39%]; 6-year-olds: 75.2% [SE = 2.41%]), though all age groups performed above chance (one-sample t-tests comparing each age group's performance to chance; all $ts > 2.13$, $ps < .05$). Finally, there was a non-significant trend for children to perform better overall on the familiar as opposed to unfamiliar relations ($F(1, 240) = 3.23$, $p = .07$).

The main effects described above were qualified by a three-way interaction between dimension, familiarity, and match type ($F(1, 240) = 8.02, p < .001, \eta^2 = .01$; Figure 3), but no other interactions reached significance ($Fs < 2.26, ps > .08$). To understand how dimension, familiarity, and match type interacted, we performed separate ANOVAs for the space-time and space-pitch conditions which each included familiarity and match type as factors. In the space-time condition, there was a main effect of match type ($F(1, 126) = 26.28, p < .001, \eta^2 = .05$), but the main effect of familiarity and the interaction term did not reach significance ($Fs < 1.9, ps > .17$). This reflects that, although children's performance did not differ significantly between the familiar space-time mapping (length-time) and the unfamiliar space-time mapping (size-time), children were better able to make both kinds of cross-dimension mappings when one of the dimensions was described by a label (mean accuracy on linguistic matching task: 78.5% [SE = 2.44%]; mean accuracy on perceptual matching task: 64.65% [SE = 2.24%]). For example, children were better at making cross-domain mappings when deciding whether a 'long' (or 'big') alien makes a long or short sound than when no labels were provided in the task. Interestingly, this suggests that labels cued children into how to perform the space-time mappings even in cases where the labels are not conventionally applied to both space and time (i.e., English 'big' and 'small' are not typically applied to time).

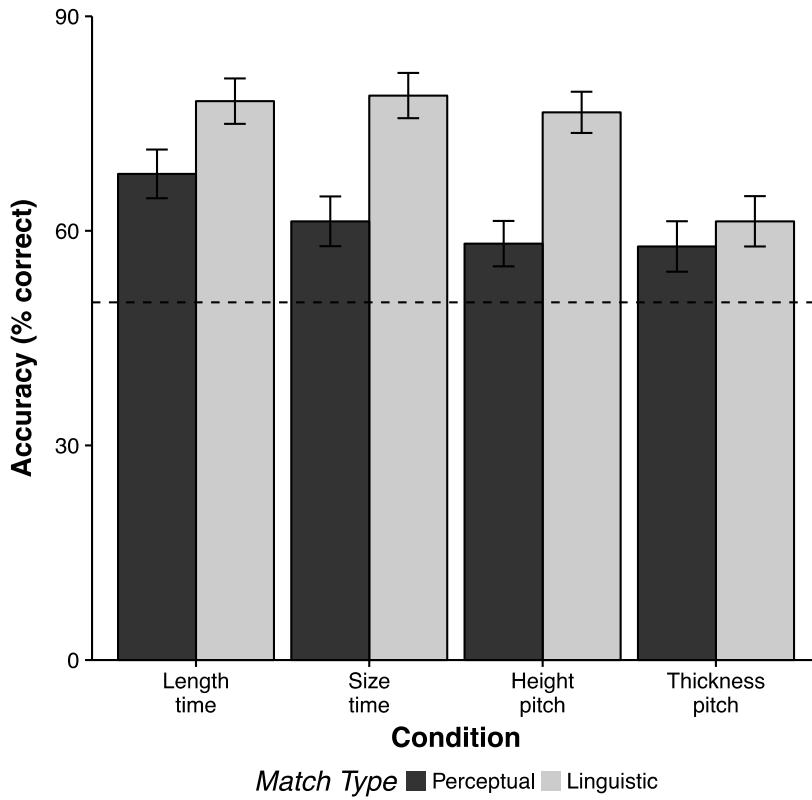


Figure 3. Children's matching performance in the perceptual and linguistic matching tasks for the four cross-domain associations. Error bars indicate +/- 1 standard error.

Meanwhile, in the space-pitch condition, the main effect of match type was significant ($F(1, 126) = 14.41, p < .001, \eta^2 = .03$), as was the interaction between match type and familiarity ($F(1, 126) = 6.64, p < .02, \eta^2 = .02$). The main effect of familiarity was also trending ($F(1, 126) = 3.28, p = .07 \eta^2 = .02$). This indicates that for perceptual matching trials, performance did not differ between the familiar height-pitch and unfamiliar thickness-pitch conditions (mean accuracy for height-pitch: 58.2% [SE = 3.18%]; thickness-pitch: 57.8% [SE = 3.53%], $t(251.25) = .08, p > .9$). For the linguistic matching trials, however, performance was significantly better in the familiar height-pitch condition compared to in the unfamiliar thickness-pitch condition (mean accuracy for height-pitch: 76.5% [SE = 2.88%]; thickness-pitch: 61.3% [SE = 2.52%]; $t(244.44)$

$= 3.35, p < .001$). Thus, for example, children were better able to make height-pitch mappings when an alien was described as ‘high’ (or ‘low’) and then had to decide which of two sounds that varied in pitch was made by the alien, compared to when no label was provided. But critically, this advantage for providing labels did not extend to ‘thick’ and ‘thin’: Children were no better at matching an alien described as ‘thick’ to one of two pitches compared to when they were matching between pictures of thick or thin aliens and pitches.

The analyses described above collapsed across the *space is source* and *space is target* match directions, because we did not have any *a priori* predictions about this factor. When we included match direction in a follow-up model, its main effect was not significant ($F(1, 240) = 1.91, p > .16$). However, it was involved in a two-way interaction between match type and match direction ($F(1, 240) = 8.99, p < .01, \eta^2 = .007$), which was subsumed by a three-way interaction between age group, match type, and match direction ($F(1, 240) = 2.91, p < .05, \eta^2 = .007$), and a four-way interaction between dimension, age group, match type, and match direction ($F(3, 240) = 2.69, p < .05, \eta^2 = .006$). These interactions are analyzed below with separate follow-up ANOVAs for the linguistic and perceptual matching tasks, which each included factors for dimension, match direction, and age group.

In the perceptual matching condition, the only significant effect was that of age group ($F(3, 248) = 6.03, p < .001, \eta^2 = .04$), demonstrating that matching performance improved with age (Figure 4). Three-year-olds performed with lower accuracy than the older age groups ($ts < -2.2, ps < .05$), and the four-year-olds performed with lower accuracy than the six-year-olds ($t = -2.66, p < .01$). In addition, three-year-olds’ performance in the perceptual matching condition was not different from chance (one-sample t-test: $t(127) = 0, p = 1$), whereas each of the older age groups performed significantly better than chance (all $ts > 3.01, ps < .005$). The main effects

of match direction and dimension, as well as all interactions, were non-significant ($Fs < 3.15$, $ps > .07$).

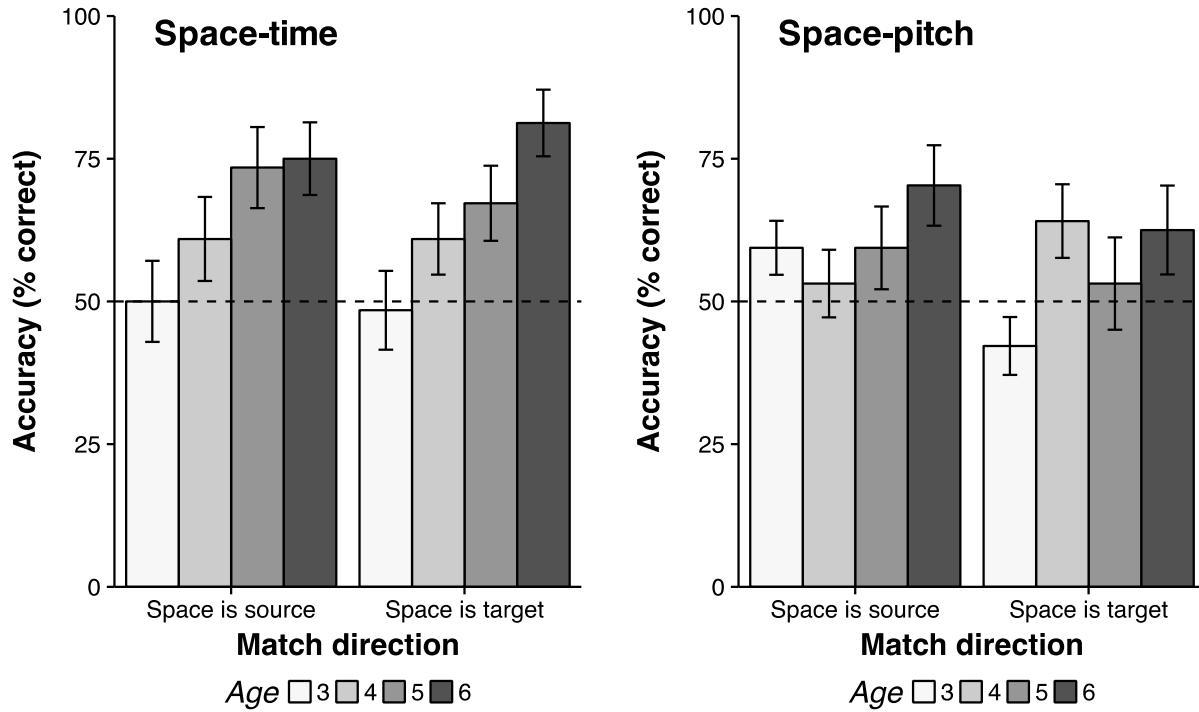


Figure 4. Children's matching performance in the perceptual matching task, separated by dimension, match direction, and age. Note that although age was the only significant factor for this task, the full interaction has been plotted for comparison with the linguistic matching task. Error bars indicate standard error.

In the linguistic matching task, the main effects of age, dimension, and match direction were all significant, as was the three-way interaction between them ($Fs > 3.22$, $ps < .01$; Figure 5). Thus, we performed separate ANOVAs for the space-time and space-pitch conditions, which each included factors for age and match direction. For space-time mappings, the main effects of age group and match direction were significant, as was the interaction between them ($Fs > 5.5$,

$ps < .005$, $\eta^2 > .04$). This reflects that there was an effect of age in the *space is source* direction (i.e., when children had to match a description of an alien's spatial attributes to the duration of an alien's sound; $F(3, 124) = 9.76, p < .001$, $\eta^2 = .19$), such that three-year-olds performed with lower accuracy than the older age groups ($ts < -3.2, ps < .005$). However, there was no effect of age in the *space is target* direction (i.e., when children had to match a description of an alien's sound to a picture of an alien; $F(3, 124) = 2.18, p = .09, \eta^2 = .05$). The effect of match direction in the space-time condition was most pronounced for the three-year-olds, who performed above chance only when matching a description of an alien sound to a picture of an alien. Older children, in contrast, were able to match labels equally well in both directions. For space-pitch mappings, the only significant effect was that of age ($F(3, 124) = 3.08, p < .05, \eta^2 = .04$), as three-year-olds performed with lower accuracy than the older age groups ($ts < -2.3, ps < .05$), and the older age groups did not differ from one another ($ts < .83, ps > .4$). These results suggest that overall, match direction did not have a strong influence on children's performance in the linguistic matching task, with the exception of three-year-olds demonstrating an advantage for matching in the *space is target* direction in the space-time condition.

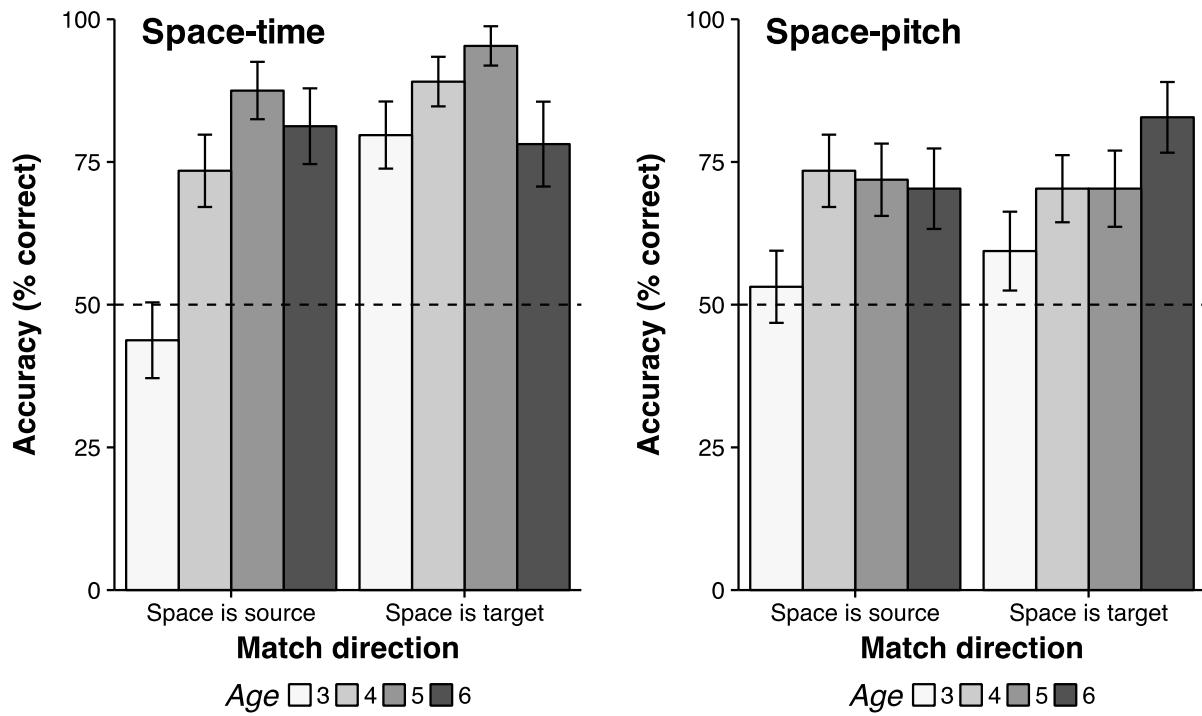


Figure 5. Children's matching performance in the linguistic matching task, separated by dimension, match direction, and age. Error bars indicate standard error.

General Discussion

The present work explored the development of cross-domain mappings between space and time and between space and pitch, and the role that spatial language may play in shaping these mappings. Three- to six-year-old English-speaking children and adults received two types of matching tasks: a perceptual matching task and a linguistic matching task (Figure 2). In the perceptual task, participants had to map between sounds that varied in duration or pitch, and aliens that varied in their size or length (for duration) and height or thickness (for pitch). In the linguistic task, participants were provided with a label that described an alien or sound in terms of its spatial attributes, auditory pitch, or temporal duration, and had to choose one of two exemplars from the other dimension to match it. Consistent with previous work (Marks, 1978;

Shayan et al., 2014; Stevens, 1957), we found that adults could easily match along multiple space-time and space-pitch relations, regardless of whether they were reflected in English metaphors, or if labels were provided for one of the dimensions. However, in contrast to previous work (Dolscheid et al., 2015; Shayan et al., 2014), we found that young children were also able to make consistent cross-domain matches that reflected both familiar and unfamiliar spatial metaphors, though children's performance was better when a label was provided for one of the dimensions in the length-time, size-time, and height-pitch conditions (Figure 3).

Critically, even though English-speaking children presumably have little experience with size metaphors for time or thickness metaphors for pitch, they were able to align these dimensions congruently with the metaphors employed in other languages. Previous results have demonstrated that preverbal infants are also sensitive to a variety of associations between space, time, and pitch (de Hevia et al., 2014; de Hevia & Spelke, 2010; Dolscheid et al., 2014; Fernández-Prieto et al., 2015; Lewkowicz & Minar, 2014; Lourenco & Longo, 2010; Mondloch & Maurer, 2004; Srinivasan & Carey, 2010; Walker et al., 2010), including the ones tested here, which suggests that these associations are either innate or constructed early in life. The present results complement these previous findings by showing that these associations are also present in preschool-aged children who are exposed only to English. These findings suggest that experience with specific spatial metaphors in language is not required for children to construct or maintain cross-domain mappings (cf. Shayan et al., 2014), and suggests a continuity with adults, who were readily able to perform cross-domain mappings in our task, and who have more generally been shown to make a variety of cross-domain mappings on explicit tasks (e.g., Dolscheid et al., 2013 Marks, 1978; Shayan et al., 2014; Stevens, 1957). However, our results remain compatible with the idea that experience with the spatial metaphors of one's language may influence the

automaticity with which cross-domain associations are accessed, which we did not measure here (Casasanto, 2008; 2010; Hespøs & Spelke, 2004).

Although our findings indicate that children can make cross-domain mappings regardless of their experience with metaphorical language, they also point to some specific ways in which labels might facilitate mappings. Recall that, although providing a label for one dimension boosted children's performance for the familiar space-time and space-pitch relations (length-duration and height-pitch, respectively), and the unfamiliar relation between size and duration, it did not do so for the unfamiliar relation between thickness and pitch. These findings suggest two ways in which labels might promote children's cross-domain mappings. First, by providing a label for one dimension, children might consider other ways in which that label can be used in their language, which could facilitate mapping to another dimension. For example, hearing an alien described as 'long' or 'high' could give children a clue that, when played two sounds that differ in duration or pitch, the sound that is 'high' goes with the 'high' alien, and the sound that is 'long' goes with the 'long' alien. By this account, the advantage for labels would depend on the labels being polysemous in the child's language – i.e., having meanings in multiple domains.

Intriguingly, however, our findings also suggest that labels might facilitate children's cross-domain mappings even when they do not have attested meanings in different domains. In particular, we found that children were better able to map between size and duration when one of these dimensions was described using the labels 'big' or 'small', compared to when these labels were not provided. For example, when an alien was described as 'big', children were more likely to indicate that of two sounds that varied in duration, the longer sound was likely made by a 'big' alien. This finding is striking, because it is unlikely that children had previously heard durations

labeled as ‘big’ or ‘small’, suggesting that labels can facilitate cross-domain mappings even when they do not have attested meanings in different domains.

We suspect that children’s size-duration mappings may have been promoted by children’s alignment of labels from different domains, according to their ordinal relations (Smith & Sera, 1992). By this account, when children heard an alien described as ‘big’ and were presented with two sounds that they might describe as ‘long’ and ‘short’ in duration, they may have mapped the word ‘big’ to the word ‘long’ because just as ‘long’ is more than ‘short’, ‘big’ is more than ‘small’. This might have helped children infer that the ‘big’ alien likely makes the long sound. This inference might not even require directly mapping between words like ‘big’ and ‘long’: Children may recognize that one presented tone is longer than another even if they do not spontaneously access the labels ‘long’ and ‘short’, and this recognition may encourage them to map a ‘big’ alien to the long tone. This account also makes sense of why providing the labels ‘thick’ or ‘thin’ did not facilitate thickness-pitch mappings in the linguistic matching task, over and above performance on the perceptual matching task (though children did perform above chance on thickness-pitch mappings in both tasks). While size and length metaphors can be readily aligned on the basis of ordinality, thickness and height metaphors seemingly ascribe opposite ordinal anchors to the spectrum of pitch: Both ‘thick’ and ‘low’ refer to low-frequency pitches, yet ‘thick’ corresponds to *more* in thickness, whereas ‘low’ corresponds to *less* in height. Furthermore, both ‘thick’ and ‘high’ correspond to the unmarked ends of their dimensions, whereas ‘thin’ and ‘low’ are the marked ends. Thus, if children attempted to align ‘thick’ or ‘thin’ aliens to ‘high’ or ‘low’ pitches based on the ordinal structure or markedness pattern of these adjectives, they would link a ‘thick’ alien with a ‘high’ sound and a ‘thin’ alien with a

‘low’ sound, i.e., exactly the reverse of how these mappings are encoded in languages that use thickness-pitch metaphors.

The idea that children can use the ordinal structure of labels from one dimension to make consistent mappings to another dimension may provide insight into why spatial labels are often used to describe other less concrete dimensions across languages (Alverson, 1994; Lakoff & Johnson, 1980; Sweetser, 1991; Xu, Malt, & Srinivasan, 2017). In particular, it may be easier for children to learn labels with spatial meanings, like ‘long’ or ‘short’, and then extend them to more abstract domains, like time, than to directly learn labels with abstract meanings. In general, words for spatial concepts may be easier to learn than words for more abstract concepts like time, because spatial labels refer to static properties of objects that can be ostensively indicated, unlike more abstract labels which are by definition less intersubjective. However, having learned the spatial meaning of a word like ‘long’, children might find it relatively easy to work out the meaning of the word when it is used in a more abstract domain, on the basis of the shared ordinal structure of these domains (e.g., such that ‘long’ refers to more in time), or through existing cross-domain associations (Srinivasan & Carey, 2010). By this account, spatial metaphors could provide a functional advantage for language learners, in scaffolding the acquisition of abstract word meanings.

Finally, we found that children’s matching performance increased with age, and that the largest change in performance occurred between ages three and four. Overall, three-year-olds performed only slightly above chance, whereas four-, five-, and six-year-olds all performed at similar levels, well above chance. This age-related increase in performance may explain why in the present study we found that children performed above chance even on unfamiliar metaphorical relations, whereas in previous work children who spoke languages with height-

pitch metaphors performed at chance with thickness-pitch metaphors (Shayan et al., 2014). In particular, in our sample, children's ages ranged from three to six years, while in Shayan et al (2014), children's ages ranged from 2.5 to four years; the wider age range of our sample could have contributed to the divergent pattern of results.

From the present data, it is difficult to determine whether the age-related jump in performance documented in our study reflects improvements in cross-domain mapping ability, or whether the demands of our matching task may have been too taxing for our youngest participants. Indeed, the only condition in which 3-year-olds performed above chance was one that required matching a description of an alien sound (e.g., a *long* sound) to a pictured alien, and this task could be solved by interpretation the adjective only in the spatial sense. Given that cross-domain associations have been demonstrated in infants (see Lourenco & Longo, 2011 for review), additional work is clearly needed to trace the development of these associations between infancy and the early preschool years. One complication in establishing continuity between infants, young children, and adults, is that different measures have been used with each population. For example, while the present study showed that, just like adults, preschoolers can make explicit mappings across domains, it is currently unknown whether the constraints on the types of cross-domain associations that adults make in more implicit tasks (e.g., Casasanto, 2008; Dolscheid et al., 2013; Srinivasan & Carey, 2010) extend to children. It remains an open question whether young children spontaneously associate space and time and space and pitch, and whether experience with particular spatial metaphors influences the automaticity with which these associations are accessed.

Conclusion

Taken together, the present findings suggest that children's cross-domain associations are not constrained to those reflected in their native language. Preschool-aged children were able to map time and pitch to spatial dimensions that both do and do not correspond to familiar spatial metaphors in the English language, which suggests that experience with specific spatial metaphors is not necessary for forming or maintaining these associations. However, we also found that labels can promote children's cross-domain mappings, both when they have familiar metaphorical uses (e.g., English 'long' denotes both length and duration), and when they do not have conventional metaphorical uses but are used to describe dimensions that share a common ordinal reference frame (e.g., size-duration but not thickness-pitch). Therefore, although the cross-domain associations present in infants may persist into childhood, language may provide additional cues for strengthening specific associations over development.

Acknowledgements

We would like to thank Gillian Schwartz, Alagia Cirolia, and Leslie Wang for their assistance with data collection. This research was supported by NIH award F32HD085736 to AS and NSF award SBE-16302040 to MS.

References

Alverson, H. (1994). Semantics and experience. Johns Hopkins Univ Pr.

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, 75(1), 1–28. [http://doi.org/https://doi.org/10.1016/S0010-0277\(99\)00073-6](http://doi.org/https://doi.org/10.1016/S0010-0277(99)00073-6)

Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, 43(1), 1–22. <http://doi.org/10.1006/cogp.2001.0748>

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1525), 1831–1840. <http://doi.org/10.1098/rstb.2009.0028>

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, 13(2), 83–91.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, 58(Suppl. 1), 63–79.

Casasanto, D. (2010). Space for thinking. *Language*, 34(3), 387–405. <http://doi.org/10.1111/j.1551-6709.2010.01094.x>

Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and time in the child's mind: Evidence for a cross-dimensional asymmetry. *Cognitive Science*, 34(3), 387–405.

de Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science*, 21(5), 653–660. <http://doi.org/10.1177/0956797610366091>

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences*. <http://doi.org/10.1073/pnas.1323628111>

de Hevia, M. D., Vanderslice, M., & Spelke, E. S. (2012). Cross-dimensional mapping of number, length and brightness by preschool children. *PLoS ONE*, 7(4), e35530. <http://doi.org/10.1371/journal.pone.0035530.g004>

Dolscheid, S., & Casasanto, D. (2015). Spatial Congruity Effects Reveal Metaphorical Thinking, not Polarity Correspondence. *Frontiers in Psychology*, 6, 1047. <http://doi.org/10.1177/0956797609354734>

Dolscheid, S., Hunnius, S., & Majid, A. (2015). When high pitches sound low: Children's acquisition of space-pitch metaphors. 37th Annual Conference of

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic Infants Are Sensitive to Space-Pitch Associations Found Across Cultures. *Psychological Science*, 25(6), 1256–1261. <http://doi.org/10.1177/0956797614528521>

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The Thickness of Musical Pitch: Psychophysical Evidence for Linguistic Relativity. *Psychological Science*, 24(5), 613–621. <http://doi.org/10.1177/0956797612457374>

Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior and Development*, 38, 77–81. <http://doi.org/10.1016/j.infbeh.2014.12.008>

Gibbs, R. W. (1994). The Poetics of Mind. Cambridge University Press.

Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, 430(6998), 453–456. <http://doi.org/10.1038/nature02634>

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago: The University of Chicago Press.

Lewkowicz, D. J., & Minar, N. J. (2014). Infants Are Not Sensitive to Synesthetic Cross-

Modality Correspondences: A Comment on Walker et al. (2010). *Psychological Science*, 25(3), 832–834. <http://doi.org/10.1177/0956797613516011>

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science*, 21(6), 873–881. <http://doi.org/10.1177/0956797610370158>

Lourenco, S. F., & Longo, M. R. (2011). Origins and development of generalized magnitude representation. In S. Dehaene & E. M. Brannon (Eds.), *Space, Time, and Number in the Brain: Searching for the Foundations of Mathematical Thought* (pp. 225–244). Elsevier. <http://doi.org/10.1016/B978-0-12-385948-8.00015-3>

Marks, L. E. (1978). The unity of the senses: Interrelations among the modalities. Academic Press.

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2), 133.

Shayan, S., Ozturk, O., Bowerman, M., & Majid, A. (2014). Spatial metaphor in language can promote the development of cross-modal mappings in children. *Developmental Science*, 17(4), 636–643. <http://doi.org/10.1111/desc.12157>

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, 24(1), 99–142.

Srinivasan, M., & Carey, S. (2010). The long and the short of it: On the nature and origin of functional overlap between representations of space and time. *Cognition*, 116(2), 217–241.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64(3), 153.

Sweetser, E. (1991). From Etymology to Pragmatics. Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9780511620904>

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P.

(2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological Science*, 21(1), 21–25. <http://doi.org/10.1177/0956797609354734>

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488.

Xu, Y., Malt, B. C., & Srinivasan, M. (2017). Evolution of word meanings through metaphorical mapping: Systematicity over the past millennium. *Cognitive Psychology*, 96, 41–53. <http://doi.org/10.1016/j.cogpsych.2017.05.005>