



Cognitive Science 36 (2012) 102–129
Copyright © 2011 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2011.01201.x

Relative Contribution of Perception/Cognition and Language on Spatial Categorization

Soonja Choi,^a Kate Hattrup^b

^a*Department of Linguistics and Asian/Middle Eastern Languages, San Diego State University*

^b*Department of Psychology, San Diego State University*

Received 13 March 2010; received in revised form 31 March 2011; accepted 31 March 2011

Abstract

This study investigated the relative contribution of perception/cognition and language-specific semantics in nonverbal categorization of spatial relations. English and Korean speakers completed a video-based similarity judgment task involving containment, support, tight fit, and loose fit. Both perception/cognition and language served as resources for categorization, and allocation between the two depended on the target relation and the features contrasted in the choices. Whereas perceptual/cognitive salience for containment and tight-fit features guided categorization in many contexts, language-specific semantics influenced categorization where the two features competed for similarity judgment and when the target relation was tight support, a domain where spatial relations are perceptually diverse. In the latter contexts, each group categorized more in line with semantics of their language, that is, containment/support for English and tight/loose fit for Korean. We conclude that language guides spatial categorization when perception/cognition alone is not sufficient. In this way, language is an integral part of our cognitive domain of space.

Keywords: Language and perception/cognition; Language and thought; Spatial categorization; Language-specific semantics; Spatial semantics; Spatial cognition

1. Introduction

A critical issue in the relation between language and thought is whether language influences our nonverbal cognition. Some studies supporting the Whorfian view have identified

Correspondence should be sent to Soonja Choi, Department of Linguistics & Asian/Middle Eastern Languages, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-7727. E-mail: schoi@mail.sdsu.edu

differences across languages in the way speakers categorize various semantic domains and have argued that these differences affect the organization of individuals' nonverbal cognition in the domains of space (e.g., Levinson, 1996; Pederson et al., 1998), time (Boroditsky, Schmidt, & Phillips, 2003), and objects (Lucy & Gaskins, 2001, 2003). The alternative "Modular" view claims that cognition is universal and is unaffected by language-specific semantics (Gennari, Sloman, Malt, & Fitch, 2002; Li & Gleitman, 2002; Malt, Sloman, Gennari, Shi, & Wang, 1999; Munnich, Landau, & Dosher, 2001; Papafragou, Massey, & Gleitman, 2002). In these studies, researchers have typically tested the relationship between language and cognition in verbal and nonverbal tasks and found *no* effects of language in nonverbal tasks. In both Whorfians' and Modularists' work, however, the question has typically been whether language does *or* does not uniformly influence nonverbal cognition.

The relationship between language and cognition may actually be more complex than has been previously suggested. We may draw on *both* linguistic and nonlinguistic resources for nonverbal organization but allocate them in different strengths depending on what is being compared and classified.¹ In fact, recent studies on infant/child spatial cognition suggest that while some categories are preverbally formed others need linguistic input to facilitate their formation. Our overall hypothesis is that language may be a critical guide in those sub-domains where perceptual/cognitive information alone is insufficient for meaningful organization or where several perceptual features compete. (See below for specific hypotheses.)

The present study explores the hypotheses by comparing English and Korean speakers' categorization of spatial relations in contexts where several features are contrasted and where these features are differentially highlighted in the two languages. Previous research on spatial categorization or Motion events generally focused on a pair of features as test items (e.g., loose vs. tight containment, or Manner vs. Path; Gennari et al., 2002; Hespos & Spelke, 2004; McDonough, Choi, & Mandler, 2003; Norbury, Waxman, & Song, 2008). The present study, in contrast, juxtaposes several features (in both the target and test items) to gain more detailed understanding about the relationship between language and cognition.

Objects can relate to one another in different ways (Fig. 1). An object can be contained, supported (on horizontal or vertical surface), attached, encircled, or covered by another, and it can fit with the other tightly or loosely. And languages differ greatly and significantly in the way they classify these relations (Bowerman, 2007; Gentner & Goldin-Meadow, 2003). Levinson, Meier, and the Language and Cognition Group (2003) examined nine unrelated languages and found an extensive diversity in how they linguistically classify static spatial relations, such as support, containment, attachment, encirclement, and contact/noncontact.

Korean and English differ in classifying dynamic spatial events, such as putting an object into/onto another, separating objects, and opening various types of containers (Bowerman & Choi, 2001). They also differ in the morphology used to categorize spatial relations, that is, "prepositions/particles" in English, "verbs" in Korean.² In English, a major distinction in spatial categorization involves whether an entity is contained (labeled *in*) or not (Fig. 2). All noncontainment relations are typically labeled *on*, whether they be horizontal support, attachment, encirclement, forming an abstract category of "support." Korean distinguishes between tight-fit and loose-fit relation. Particularly, the verb *kkita* "tight fit or interlock" expresses "tight-fit" relation, whether it involves containment or support. When the relation

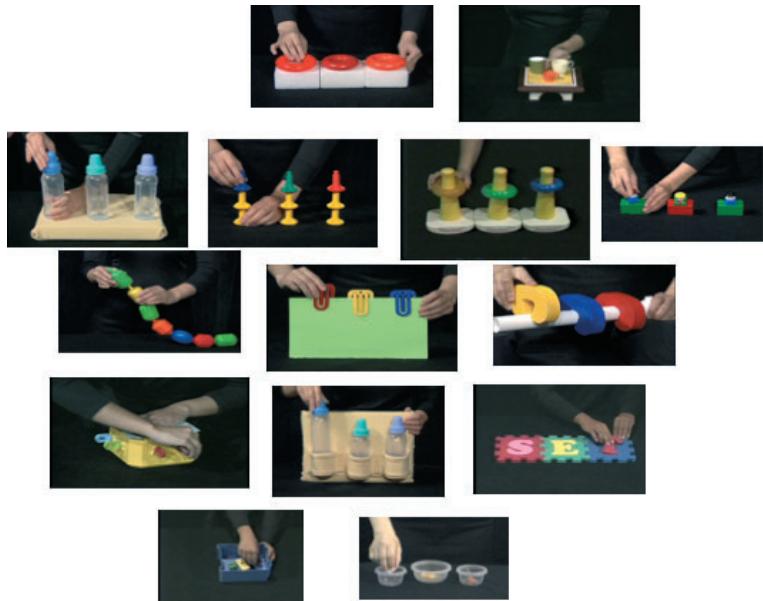
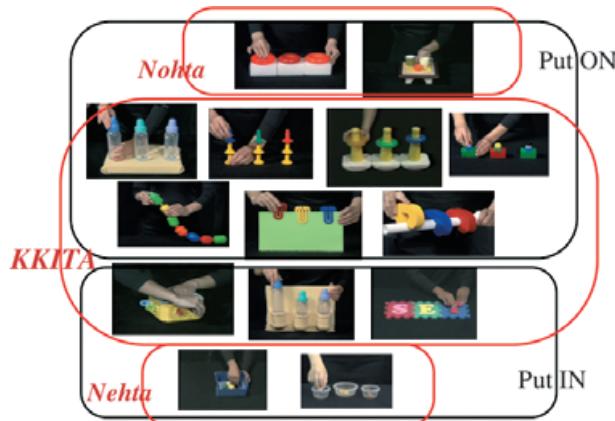


Fig. 1. Diversity of spatial relations.

Fig. 2. Semantic categorization of containment and support in English (*put in* and *put on*) and Korean (*kkita*, *nehta*, and *nohta*).

does not involve tight fit, a distinction is made between loose containment (*nehta*) and loose support (*nohta*). But here again, Korean differs from English: The category of *nehta* (a word generally referring to loose containment) includes loose encirclement as well, for example, a big ring *on* a thin pole. Thus, the division between containment and support is again blurred in the two loose-fit categories in Korean (*nehta* and *nohta*).³

Given the different semantic categories in the two languages, two central questions arise about nonverbal categorization for space: (1) What are the perceptual/conceptual bases for nonverbal categorization in adult speakers? (2) To what extent, if any, do the language-specific semantics guide nonverbal spatial categorization?

As mentioned above, clues for the answers come from recent developmental studies on spatial categorization as infants start categorizing and generalizing spatial relations from the preverbal stage and continue to do so as they learn their first language. Thus, developmental patterns of spatial categorization from infancy to early stages of language learning would give us some insight into what is perceptually/cognitively universal in spatial categorization and whether and how much the categorization is influenced by language-specific semantics. In the next section, we briefly review studies of spatial cognition involving containment, support, tight, and loose-fit relations in Korean- and English-learning infants and children. These studies will serve as bases for our specific hypotheses that will follow.

1.1. Developmental studies on spatial categorization from infancy to early childhood

Infants explore the physical properties of spatial relations (e.g., containment, support, tight fit) virtually from the beginning of life. For example, 2.5-month-old infants can already recognize that containment involves an opening (for a Figure [= moving] object to go in) and that the object contained moves with the container (Hespos & Baillargeon, 2001). Forming an abstract *category* of containment, however, seems to take a little more time: Casasola, Cohen, and Chiarello (2003) showed that infants can generalize spatial relations with novel objects (novel both in the Figure object and the container) at 6 months of age based on the containment feature and thus form an abstract category of containment. Thus, containment is a perceptual/conceptual category developed during the first year of life.

Preverbal infants are also sensitive to certain types of tight fit and can distinguish it from its loose-fit counterpart. In particular, within the containment relation, they can detect degree of fit, that is, tight- versus loose-fit containment (McDonough et al., 2003). Using the preferential looking paradigm with scenes depicting a variety of objects, McDonough et al. (2003) showed that from 9 months of age, preverbal infants (regardless of language environments) can categorize spatial events on the basis of tight versus loose feature within the containment relation (i.e., tight containment vs. loose containment).

Detection of the tight-fit feature can start even earlier. Hespos and her colleagues (Hespos & Piccin, 2009; Hespos & Spelke, 2004), using the habituation paradigm, have shown that 5-month-olds can distinguish between tight and loose fit within the covering relation and can generalize its tight-fit feature to tight-fit containment (Hespos & Piccin, 2009). It should be noted, however, that in these studies the same type of objects (e.g., cylindrical objects) was used for both habituation and test scenes. With novel objects, however, there is evidence that infants show difficulty in generalizing all types of tight fit (that include, e.g., tight-fit containment, tight-fit attachment, tight-fit covering, tight-fit encirclement) as a single category. Casasola and Cohen (2002) report that 10-month-olds do not generalize tight fit as a category, and even at 18 months, infants do not fully generalize tight fit to novel objects. This may be because young children perceive various types of tight support—tight

encirclement (e.g., putting a ring tightly onto pole), tight attachment (e.g., putting a Lego piece tightly on another), tight covering (e.g., putting a cap tightly on bottle; see Fig. 3)—to be all distinct relations (Choi & Casasola, 2005).

Linguistic input is reported to play a crucial role in the categorization of tight fit in young children. In Casasola, Bhagwat, and Burke's (2009) study, 18-month-olds (learning English as first language) could form an abstract category of tight-fit relation with linguistic training (hearing a novel label during habituation) but could *not* do so while viewing the habituation events in silence. (In both conditions, the test trials were conducted in silence.) The study suggests that the formation of a tight fit category at the perceptual/conceptual level needs guidance from language.

As for the “support” relation, while the physical properties of loose support are explored from 4 months of age (Needham & Baillargeon, 1993), an abstract category of support—a category that includes both loose and tight support—is reported to develop relatively late (Casasola, 2005b; Casasola & Cohen, 2002). This may be due in part to infants’ perception that loose support (e.g., putting a cup on the table, see Fig. 4) is distinct from tight support (e.g., putting a Lego piece on another, putting a ring tightly onto a pole; Choi & Casasola, 2005). Also, the process of representing all types of support relation as a unified category appears to unfold in a piecemeal fashion: Casasola (2005a) reports that when English-learning infants begin to generalize the support relation at around 14 months, they can do so only with a small number of exemplars (e.g., with two but not with six exemplars). Furthermore, even at 18 months, infants need linguistic input (e.g., hearing the word “on” during the habituation phase) to generalize the support relation to novel objects (Casasola, 2005b; see also Casasola & Bhagwat, 2007). These findings suggest that it is language-specific semantics that unite different types of support (loose support, tight encirclement, tight attachment, tight covering) to a single category at the conceptual level.

To summarize, preverbal infants (regardless of the ambient language) can form abstract categories of containment, specific types of tight fit (e.g., distinction of tight containment from loose containment), and loose support, suggesting that these categories are driven by

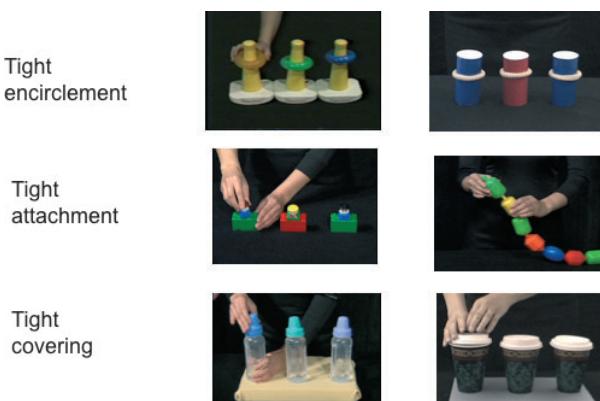


Fig. 3. Different types of tight support.



Fig. 4. Examples of loose support.

our nonlinguistic perceptual/conceptual ability. The ease with which these categories are formed is probably due to a high degree of perceptual coherence within each category. In contrast, infants have difficulty in forming abstract categories of tight fit (that include all types of tight containment and tight support), support (that include both loose and tight support), and tight support (that include tight encirclement, tight attachment, and tight covering). Note that the types of spatial configuration among the members of these categories are quite diverse. Studies have shown that infants develop these latter categories in a piecemeal fashion and that they need linguistic guidance to form these categories.

Children are sensitive to linguistic input from very early on. Using the preferential paradigm, Choi, McDonough, Bowerman, and Mandler (1999) demonstrated that by as early as 17 months, English learners and Korean learners comprehend the terms “in” and “kkita,” respectively, in a language-appropriate way. Such early acquisition of spatial terms is no doubt founded on perceptual salience for containment and certain types of tight fit, but it also points to infant’s sensitivity to language-specific input. Furthermore, there is evidence that over time, language-specific meanings affect, at least partially, children’s nonverbal categorization: from 29 months of age, the ability to perceive tight and loose containment as distinct categories weakened in English learners, whereas the categorization continued in Korean learners (Choi, 2006a). The change in English learners’ sensitivity resided mostly in their inability to detect looseness, as they were still able to detect the tight-fit feature,⁴ indicating that whereas language-specific semantics influence nonverbal classification, it does so only partially and certain perceptual salience (i.e., tight-fit feature within the containment relation in this case) persists regardless of language (Choi, 2006a).

The asymmetry between tight- and loose-fit features seems to continue through the adult years. Norbury et al. (2008) also found an asymmetry between tight-fit and loose-fit features in adult speakers of English: They could pick out the tight-fit feature better than the loose-fit feature in a similarity judgment test. But unlike Choi’s (2006a) finding, Norbury et al. (2008) found that the asymmetry is present in Korean speakers as well, resulting in no cross-linguistic difference. The lack of a language effect in Norbury et al.’s study, however, may be due to the nature of the task and the scope of the stimuli: Participants rated the degree of similarity between a test relation and the target relation using a 10-point scale where individual differences are possible in the rating criteria. Also, for all target relations (tight-in, loose-in, tight-on, or loose-on), the two test relations were constant (i.e., always tight-in vs. loose-in) and all relations (both for the target and the test) were depicted with the same types of Figure and Ground objects; thus, the range of exemplars was limited. Because spatial categories are “relational” in nature (i.e., generalized across a number of objects), to understand the interaction between perception/cognition and language, we need to test categorization with a variety of objects, a design that the present study takes.

1.2. The present study and hypotheses

We hypothesize that there are two sources for nonverbal spatial categorization: universal perception/cognition and language-specific semantics. Specifically, we hypothesize that while our universal perception/cognition is an important and indispensable resource for classification, language also plays an important role by guiding our spatial cognition in areas where the relations are perceptually diverse and where several perceptually salient features compete for classification. Based on the spatial semantics of English and Korean and the developmental literature just reviewed, we make the following hypotheses and predictions about Containment, Support, Tight-fit, and Loose-fit relations. The hypotheses for containment and support are presented separately as the literature suggests that the two relations interact with language in different ways.

Hypothesis 1. Containment and tight/loose features

Hypothesis 1a : Containment and degree of fit within the containment relation (i.e., tight containment vs. loose containment) are features that all speakers, regardless of language, can perceive and utilize for categorization. Thus, when either containment or tight fit is the only differential feature for categorization, both language groups will categorize similarly basing their choices on containment or tight-fit feature, respectively.

Hypothesis 1b : When the two features—containment and tight fit—compete for categorization, each language group will categorize more in line with language-specific semantics. In this context, English speakers will categorize more on the basis of containment and Korean speakers will do so more on the basis of tight/loose-fit features.

Hypothesis 2. Support and tight/loose features

Hypothesis 2a : Loose support will be perceptually distinguished from tight support (cf. Choi & Casasola, 2005). However, English speakers may perceive the two types of support to be relatively more similar to each other than Korean speakers (i.e., they pick out the support feature more than Korean speakers). (The latter hypothesis is based on the finding that from an early stage language facilitates the formation of a “support” category that includes both loose and tight support [Casasola, 2005b].)

Hypothesis 2b : Categorization of tight support will be guided by language: English speakers will pick out the support feature for categorization more often than will Korean speakers, and Korean speakers will pick out the tight-fit feature more often than will English speakers.

Overall, we hypothesize that degrees of perceptual coherence are different for containment and support. While the containment relation forms a perceptually/conceptually coherent category, the support relation does not. Language (e.g., “on” in English) is needed to put all types of support relations together as a single category at the perceptual/conceptual level. Also, while detecting the property of tight fit within containment is part of our perceptual ability, detecting tight fit from various types of tight support events will need linguistic guidance (e.g., “kkita” in Korean).

We test these hypotheses with a forced-choice similarity judgment task with video clips involving dynamic actions of tight containment, loose containment, tight support, and loose support. If our hypotheses hold, our results would demonstrate that language and cognition interact in the organization of spatial relations. It would also pinpoint the nature of the interaction and the contexts in which each type of resource is more strongly used during an online nonverbal categorization task.

2. Method

2.1. Participants

Two hundred thirty-one English speaking undergraduates, recruited from universities in Southern California and 229 Korean students, recruited and tested in South Korea, participated in the study for either course credit (English speakers) or a small gift (Korean speakers). All participants reported being monolingual with minimal foreign language exposure (restricted to classroom experience in Korea). Data from eight English speakers and four Korean speakers were excluded because of computer failures during the experiment. The final analysis included 223 English speakers (118 females and 105 males) and 225 Korean speakers (115 females and 110 males) who were randomly assigned to one of the three groups and to one of the three experimental conditions within each group, as described below. This resulted in having between 23 and 27 participants in each condition in each group.

2.2. Stimuli, design, and procedure

A computer-based similarity judgment task was administered to participants. For each trial, a *target* video depicting an actor performing a particular spatial relation (e.g., putting *x* in a container, see Appendix) appeared on the screen (5 s), then disappeared and was followed by a screen depicting two *choice* videos shown side by side (Fig. 5a and b). Participants were instructed to choose the video that was more similar to the target, by clicking a check box provided below each choice scene (Fig. 5b). All choice videos ran for 5 s, then paused on the last frame until a choice was made. After the participants selected a video, a new trial began. Choices were recorded by the computer.

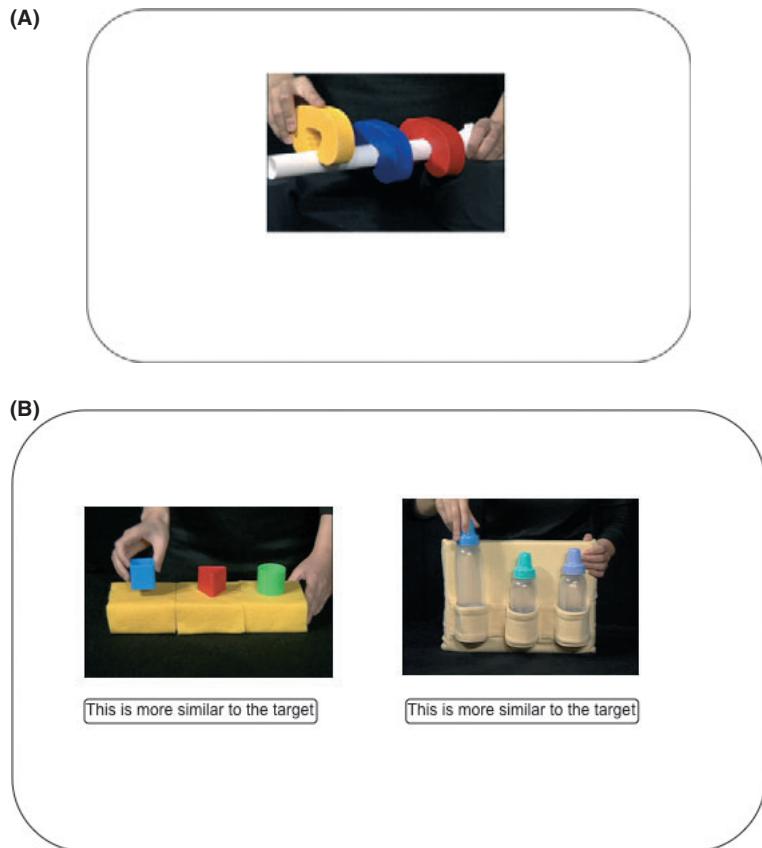


Fig. 5. (A) A still frame of a target video. (B) Still frames of a pair of choice videos with a click box below each video.

2.2.1. Test stimuli: Design of triad types

Four types of spatial relations were tested: (1) tight containment (tight-in), (2) loose containment (loose-in), (3) tight support (tight-on), and (4) loose support (loose-on). The target video depicted one of the four types, and the two choice videos could be any two of the four relations. A target relation followed by two choice relations together form a “triad.” Of possible 24 triad types, a total of 18 were tested (see Table 1).⁵

For each triad type, six trials with different exemplars were constructed (Table 2), totaling 108 trials. There were three target videos per triad type, each used in two of the six trials. However, because the pair of choice videos differed in each trial, each trial consisted of a unique combination of a target and two choice videos. The same choice exemplar could be used twice. A target video was never used as a choice video and vice versa. The sides of the screen where the two choices were presented were counter-balanced.

All the exemplars depicted perceptual features appropriate for the linguistic labels relevant to the present study (i.e., *in*, *on* in English, *kkita* [‘tight fit’], *nehta* [‘loose

Table 1
Triad types tested in the study^a

	Target	Choice 1	Choice 2
1	Loose IN	<i>tight in</i>	<i>tight on</i>
2	Loose IN	<i>loose in</i>	<i>loose on</i>
3	Loose IN	<i>tight in</i>	<i>loose in</i>
4	Loose IN	<i>tight in</i>	<i>loose on</i>
5	Tight IN	<i>tight in</i>	<i>tight on</i>
6	Tight IN	<i>loose in</i>	<i>loose on</i>
7	Tight IN	<i>tight in</i>	<i>loose in</i>
8	Tight IN	<i>tight on</i>	<i>loose on</i>
9	Tight IN	<i>tight on</i>	<i>loose in</i>
10	Loose ON	<i>tight on</i>	<i>loose on</i>
11	Loose ON	<i>loose on</i>	<i>loose in</i>
12	Loose ON	<i>tight in</i>	<i>tight on</i>
13	Loose ON	<i>loose in</i>	<i>tight on</i>
14	Tight ON	<i>tight on</i>	<i>loose on</i>
15	Tight ON	<i>tight in</i>	<i>loose in</i>
16	Tight ON	<i>tight on</i>	<i>tight in</i>
17	Tight ON	<i>loose on</i>	<i>loose in</i>
18	Tight ON	<i>tight in</i>	<i>loose on</i>

Notes. ^aExperimental Group A was tested on Triad types 1, 2, 5, 7, 14, and 17; Group B on Triad types 3, 4, 10, 13, 15, and 16; and Group C on Triad types 6, 8, 9, 11, 12, and 18. Thus, each group was tested on six different sets of triad types.

Table 2
Design of Triad Type: An example with Loose-IN as target, and *tight-in* and *tight-on* as choices (Triad Type 1)

Triad Type 1	Target (Loose-IN)	Choice on Left Screen	Choice on Right Screen
Trial 1	Rings in a basket	<i>Tight-in:</i> books in matching covers	<i>Tight-on:</i> wheels on poles
Trial 2	Rings in a basket	<i>Tight-on:</i> tops on bottles	<i>Tight-in:</i> lego men in lego cars
Trial 3	Shapes in boxes	<i>Tight-in:</i> bottles in matching pouches	<i>Tight-on:</i> lego men on lego blocks
Trial 4	Shapes in boxes	<i>Tight-on:</i> toy rings on poles	<i>Tight-in:</i> candles in slots
Trial 5	Candies in a bowl	<i>Tight-in:</i> wood posts in matching holes of a wood board	<i>Tight-on:</i> pencil erasers on tip of pencils
Trial 6	Candies in a bowl	<i>Tight-on:</i> thimbles on fingers	<i>Tight-in:</i> straws in matching holes of the lid into cups

containment & loose encirclement''], and *nohta* ['loose support'] in Korean). Thus, all the exemplars depicting tight containment and tight support had a tight-fit feature appropriate for the semantics of *kkita* in Korean, and *in* and *on* in English, respectively. All the exemplars of loose containment showed small objects (e.g., pegs, candies) being put into a container (*in* in English and *nehta* in Korean), and all exemplars of loose support showed small objects being put onto a flat horizontal surface (*on* in English and *nohta* in Korean). Descriptions of these scenes were sampled from 12 native monolingual speakers of each language.

The speakers' description contained verb phrases in both languages: "verb + spatial preposition/particle (+noun phrase)" in English, and "(noun phrase+) verb" in Korean.⁶ The speakers of each language used the relevant linguistic labels predominantly and appropriately in their descriptions: Overall, English speakers used the preposition *in* for the containment events in 85% of their descriptions and *on* for the support events in 76% of their descriptions. (The English speakers expressed tight fit only 0.05% in their descriptions with verbs *fit* or *snap*.) The Korean speakers used the verb *kkita* for the tight-fit events 73%, *nehta* for the loose containment events 80%, and *nohta* for the loose support events 76% in their descriptions. Thus, overall, the video stimuli of the present study reliably exemplified the meanings of the spatial words investigated in the present study.

Objects in the videos were of different colors and different shapes (Table 3), as the goal of the study was to examine a broad array of objects to assess patterns of categorization based on relational features regardless of object properties. In a given trial, a particular object could appear only in *one* of the three scenes of the triad, eliminating the possibility for judging the similarity based on the sameness of object.

With 108 test trials, the possibility of participant fatigue and strategy building was high. Thus, we divided the 18 triad types into three Groups (A, B, C), resulting in each Group being tested on different sets of 36 trials (6 triad types × 6 trials; see note in Table 1 for Group assignment). The 36 trials were presented in a random sequence.

2.2.2. Procedure and familiarization trials

Participants were tested individually in a quiet testing room. First, the participant completed a language background questionnaire. Then, the experimenter read the task instructions to the participant (see Appendix).

Table 3
Examples of target and choice scenes^a

Target scenes

Tight-IN: shapes in matching holes; puzzle pieces in slots; keys in locks; corks in bottles

Loose-IN: plastic shapes in container; pencils in cup; plastic rings in basket; candies in bowl

Tight-ON: plastic paperclips on paper; foam door clamps on a cylindrical bar; pentops on pen; legos on lego baseboard

Loose-ON: teacups on table; wooden blocks on block; plastic toy rings on plastic flat block

Choice scenes

tight-in: legomen in lego cars; wooden posts in matching holes in wood block; books into their covers; baby bottles in matching pouches

loose-in: blocks in rectangular container; books in paper bag; books in plastic bins; toys in basket

tight-on: tops on baby bottles; pencil erasers on pencil ends; top part of wooden russian doll onto bottom part; rings on fingers

loose-on: crayons on paper; figurines on miniature stairs; thimbles on back surface of hand (horizontal position); photos on photo album sheet

Notes. ^aThe stimuli objects were of various colors and shapes. In a given trial, a particular type of object (e.g., Lego pieces) appeared only in *one* of the three scenes of the triad. Thus, a similarity judgment could not be based on the sameness of object.

The participant completed two familiarization trials that depicted actions that did not contain spatial relations. In the familiarization trials, one of the choices was identical to the target to encourage the participant to choose the most similar alternate in the test trials. Participants could repeat the familiarization trials but none requested to do so. Immediately after familiarization, the test trials began. The task took approximately 15 min to complete.

2.2.3. Scoring

We were interested in documenting similarities and differences between the two language groups in their choices in a particular triad type. Depending on the triad type, the features (containment, support, tight, loose) compared in the target and choices differed: the choices could contrast tight versus loose features, containment versus support, or tight/loose versus containment/support. Thus, the assignment of which spatial relation (between the two choices) receives the score of “one” (or “zero”) cannot be constant and is not related to “correctness.” However, for consistency, we used the following general procedure: (a) Score of 1 on the basis of the similarity of containment/support when the choices contrast containment versus support. (b) Score of 1 on the basis of the similarity of tightness/looseness when the choices contrast tight versus loose. Thus, overall, higher scores are received across a set of six trials when participants’ choices match the target on one of the salient dimensions, tightness/looseness, or containment/support. If language and perception/cognition contribute differentially to nonverbal spatial categorization depending on the spatial relation and the categorization context (i.e., what the target relation is and which feature is being contrasted in the two choices) as hypothesized above, we would expect an interaction between Language and Triad type in our results.

2.2.4. Experimental conditions

The similarity judgment was performed nonverbally, as our goal was to identify the interaction between perception/cognition and language in nonverbal spatial categorization. To assess the degree to which participants might be engaged in verbal thinking during the experiment, and whether this affected our results, we conducted the experiment in three conditions: (a) silent, (b) shadowing, and (c) linguistic (adapting the methodology used in Gennari et al., 2002). In the silent condition, participants performed the judgment task in silence. In the shadowing condition, participants performed the task while at the same time repeating a string of nonsensical three Consonant–Vowel syllables heard from a CD player, which could minimize possible verbal thinking. In the linguistic condition, participants were first shown all and *only* the target scenes one by one (18 scenes in total: 3 target scenes × 6 triad types) and were asked to describe them. To reduce an immediate impact of linguistic descriptions of the target scenes on the similarity judgment, the description task was followed by an interval of 5–10 min during which time the participants in this condition solved a set of nine arithmetic problems in a written form. After this interval, they performed the similarity judgment task in silence.

2.3. Post-experiment survey

At the conclusion of the experiment, participants were asked two questions about the degree to which they were engaged in verbal thinking during the task: (a) *Did you think of specific words during the task*, and if “yes,” (b) *what were the words?* Although a post-experiment interview cannot assess accurately participants’ thought process during the experiment, it does give us some indication about the degree of linguistic interference during a nonverbal task.

Overall, participants reported an average of 1.7 words ($M = 1.33$ [$SD = 1.49$], shadowing; $M = 1.78$ [$SD = 1.47$], silent; $M = 2.12$ [$SD = 1.83$], linguistic). Of this number, spatial words were mentioned for only 0.7 word on average ($M = 0.37$ [$SD = 0.89$], shadowing; $M = 0.78$ [$SD = 1.14$], silent; $M = 0.98$ [$SD = 1.41$], linguistic). (The other words reported included words about object properties [e.g., color and shape] and object location [e.g., *left*, *right*.]) Although participants in the linguistic condition reported more words overall and a larger proportion of spatial words than did participants in the shadowing condition [total words: $t(416) = 4.07$, $p < .01$ and proportion of spatial words: $t(307) = 3.26$, $p < .01$], the difference was less than a word ($M = 2.12$ vs. $M = 1.33$ for total words and $M = 0.98$ vs. $M = 0.37$ for spatial words in linguistic vs. shadowing condition, respectively), thus quite small. None of the other pairwise differences were statistically significant when adjusted using the Bonferroni correction. Overall then, the self-report data suggest that the participants across the three conditions engaged in very little verbal thinking, particularly little verbal encoding of spatial words, during the experiment.

3. Results

Participants’ choice scores were analyzed with a $2 \times 18(3) \times 6(18) \times 3$ (Language by Triad type within Group by Example within Type by Condition) binomial chi-square analysis, implemented with Generalized Estimating Equations in SPSS, that treated the six Example triads as events/trials within levels of Type. Triad types represented a within-subjects factor nested in Groups (A, B, C), and Language (English, Korean) and Condition (silent, shadowing, linguistic) were both between-subjects factors. Main effects were observed for each of the factors in the design [Wald $\chi^2(2) = 135.01$, $p < .01$, Group; Wald $\chi^2(1) = 18.37$, $p < .05$, Language; Wald $\chi^2(15) = 735.182$, $p < .01$, Triad Type; Wald $\chi^2(1) = 10.57$, $p < .05$, Condition]. Furthermore, as expected, a significant interaction was observed between Language and Triad type [Wald $\chi^2(17) = 94.24$, $p < .01$], indicating that differences between English and Korean speakers in their choices depended on the Triad type. Note that each Triad type has a unique combination of a target and two choice relations. The main effect of Group is the result of each group being tested on a unique set of triad types (Table 1), and therefore, differences across groups were expected. The Group effect simply means that different sets of Triad types resulted in different scores. Similarly, the main effect of Triad type reflects the expected effects of Triad type differences on participants’ choices, and the Language effect means that Korean- and English-speaking

participants differed in their choices overall. Importantly, however, these two main effects (i.e., Triad type and Language) were qualified by a significant interaction, indicating that language contributed differentially in different Triad types.

Condition showed a main effect, but it did not interact with any other factor. The main effect of Condition was the result of a relatively lower mean score across triad types in the shadowing condition ($M = 3.82$, $SD = 1.28$) compared to the other silent and linguistic conditions ($M = 3.95$, $SD = 1.23$; $M = 4.1$, $SD = 1.24$, respectively). The overall lower mean scores in the shadowing condition reflect a depressed pattern of choices probably as a result of the cognitive interference associated with repeating nonsense syllables during the experiment. (See Gennari et al. [2002] and Trueswell and Papafragou [2010] for similar results for shadowing/interference conditions.)

Condition did not interact with any other factor. In particular, Condition did not interact with Language [Wald $\chi^2(2) = 1.50$, $p = .472$] nor did it interact with Triad Type [Wald $\chi^2(34) = 39.97$, $p = .222$]. A three-way interaction, Condition by Language by Triad Type also did not occur either [Wald $\chi^2(2) = 44.02$, $p = .117$]. This means that the observed pattern of interaction between Language and Triad Type was similar and present across the three conditions. We interpret this to mean that the participants in all three conditions performed the task in a similar manner. Given a convergence among recent findings that nonverbal tasks in silent and shadowing conditions are indeed performed with little verbal encoding (Gennari et al., 2002) especially when the interfering material is linguistic (Trueswell & Papafragou, 2010), such as nonsense syllables or counting, we interpret the similar pattern of choices across the three conditions to mean that the participants in all three conditions performed the task nonverbally to a similar degree.⁷ This conclusion is also supported by the results of the self-report in the present study that the participants engaged in little linguistic encoding during the task. At the same time, it is also possible that linguistic processing was used by participants in this study and we were just unable to detect it. Future research might be designed to address this possibility.

We now focus on the interaction between Language and Triad Type observed across the three conditions and report follow-up tests that compared languages at each level of Triad type using chi-squared tests. The significant interaction between Language and Triad type (henceforth TT) indicates that language contributes differentially in different TT contexts. To understand the locus of this interaction, follow-up comparisons of each TT between languages were tested using chi-square analyses, treating the six triad exemplars as events/trials in each analysis. To control familywise type I error, the Bonferroni adjustment was used, so that comparisons at $p < .0028$ (.05/18) were considered statistically significant. Table 4 presents the mean values and standard deviations of participants' choices across exemplars for each Triad Type for each Language group. The Wald chi-square for each Triad Type is presented, as is the R^2 for the Language contrast.⁸ Among the 18 TTs, choice scores differed significantly between English and Korean for five TTs (TTs 8, 9, 15, 17 and 18), and one TT (TT 16) closely approached significance. We report these differences as well as the similarities by evaluating each hypothesis. (The order of the TTs in Table 4 differs from Table 1, and it is organized according to the sequence of the hypotheses listed above.)

Table 4

Triad types, their average choice scores by language, chi-square results, and regression analysis results for the language contrast

Triad Type	Target: Choices	English <i>M</i> (<i>SD</i>)	Korean <i>M</i> (<i>SD</i>)	Wald χ^2	<i>p</i> -Value	<i>R</i> ²
Target: tight or loose containment						
TT 1	T-IN^a: l-in^b versus l-on	4.90 (1.05)	4.68 (1.20)	1.35	.246	.009
TT 2	L-IN: t-in versus t-on	4.78 (1.15)	4.58 (1.17)	1.10	.298	.007
TT 3	T-IN: t-in versus t-on	4.01 (1.05)	3.91 (1.10)	0.36	.546	.002
TT 4	L-IN: l-in versus l-on	4.77 (1.17)	4.48 (1.29)	2.21	.147	.013
TT 5	L-IN: l-in versus t-in	3.89 (1.31)	4.24 (1.11)	2.99	.084	.021
TT 6	T-IN: t-in versus l-in	4.63 (1.16)	5.05 (1.07)	5.64	.018	.035
TT 7	T-IN: t-on versus l-on	4.99 (0.89)	4.97 (1.13)	0.01	.938	.000
TT 8 ^c	L-IN: t-in versus l-on	3.63 (1.48)	2.83 (1.41)	11.07	<.001	.071
TT 9 ^c	T-IN: l-in versus t-on	2.90 (1.36)	1.72 (1.42)	22.25	<.001	.143
Target: loose support						
TT 10	L-ON: l-on versus t-on	4.30 (1.84)	4.60 (1.24)	3.22	.073	.021
TT 11	L-ON: t-on versus t-in	3.00 (1.39)	2.64 (1.09)	3.05	.081	.021
TT 12	L-ON: t-on versus l-in	2.46 (1.31)	2.07 (1.39)	3.02	.082	.021
TT 13	L-ON: l-on versus l-in	3.61 (1.36)	3.46 (1.36)	0.47	.497	.003
Target: tight support						
TT 14	T-ON: t-on versus l-on	4.43 (1.04)	4.60 (1.18)	0.77	.380	.006
TT 15 ^c	T-ON: t-in versus l-in	4.41 (1.29)	5.10 (1.09)	11.42	.001	.075
TT 16 ^d	T-ON: t-in versus l-on	4.40 (1.20)	4.96 (1.02)	8.97	.003	.060
TT 17 ^c	T-ON: t-on versus t-in	3.74 (1.00)	3.08 (1.20)	12.80	<.001	.082
TT 18 ^c	T-ON: l-on versus l-in	3.77 (1.34)	2.83 (1.39)	17.60	<.001	.105

Notes. ^aL-IN: loose containment; T-IN: tight containment; L-ON: loose support; T-ON: tight support.

^bBetween the two choices, the bold-faced relation received a score of 1 when it was chosen. There were a total of six trials for each Triad Type. Therefore, the maximum score a participant would get for a given triad type was 6.

^cCross-linguistic difference is significant, $p < .0028$.

^dCross-linguistic difference closely approached significance.

3.1. Containment as target relation

Hypothesis 1 predicted that containment and tight fit (within containment) are relations that all speakers can perceive and utilize for categorization (1a), but that when the two features compete for categorization, English and Korean speakers will diverge, each categorizing more in line with language-specific semantics, that is, containment for English and tight fit for Korean (1b). This hypothesis was evaluated in TTs 1 through 9 where containment (tight or loose) was the target relation and the choices contrasted containment with noncontainment (= support; TTs 1–4), or tight fit with loose fit (TTs 5–7), or containment with tight/loose fit (TTs 8–9). Among these TTs, Hypothesis 1 would predict cross-linguistic differences in only TT 8 and TT 9 where containment is pitted against tight/loose fit for categorization. Indeed, a significant cross-linguistic difference was obtained only for TTs 8 and 9. All other TTs (1 through 7) yielded cross-linguistic similarity, although TTs 5 and 6

showed some tendency for cross-linguistic difference. We examine each part of the hypothesis, 1a and 1b, in more detail.

Hypothesis 1a predicted that containment and tight/loose features within the containment relation are perceptually/conceptually salient features for speakers of all languages. This hypothesis predicts that when either containment or tight fit is the only differential feature for categorization, both language groups will categorize similarly basing their choices on containment or tight/loose fit feature, respectively.

This hypothesis was tested across TTs 1–7 in which the two choices differed by only one feature, either containment (TTs 1–4) or tight fit (TTs 5–7). As shown in Table 4, these TTs did not yield any cross-linguistic differences, suggesting that for both groups, containment and tight fit served as strong perceptual cues for categorization. More specifically, TTs 1–4 tested whether containment is a perceptually salient feature. When the target was containment (tight or loose) and the choices contrasted containment versus support (with the tight/loose feature held constant in the choices), both groups chose on the basis of the containment feature. The salience of the containment feature is particularly demonstrated by the performance of Korean speakers in TTs 1, 2, and 3. If the Korean speakers were influenced only by the semantics of their language, their choices in these triads would have shown a lack of preference between the two choices (because both choices are either equally different from the target [TT1 and 2] or in the same semantic category as the target [TT3]). However, in all three TTs, the Korean participants consistently preferred choosing containment as more similar to the target.

Along the same line, the salience of the tight-fit feature is demonstrated in TTs 5, 6, and 7 where the tight versus loose was contrasted with the containment (TTs 5 and 6) or support feature (TT 7) held constant in the choices. In these TTs, both groups matched the target on the basis of the tight/loose feature. Here, the performance by English speakers is particularly revealing. In TTs 5 and 6, the two choices, *tight-in* and *loose-in*, are members of the same semantic category as the target relation (Loose-in) in English and in TT 7, *tight-on* and *loose-on* are both semantically distinct from the target Tight-IN. If English speakers were only guided by the spatial semantics of English, they would not show any preference between the two choices in these triads. However, they consistently chose on the basis of tight fit (or loose fit), showing that they pick out the tight-fit (or loose-fit) feature from a containment relation. In particular, in TT 7 (where both choices involved support [a different relation from the target containment relation in English semantics] and contrasted tight vs. loose support), English speakers matched the target with the tight-fit feature to almost the same degree as Korean speakers.

The marginal cross-linguistic differences in TTs 5 and 6 should be pointed out, however. Although these TTs did not yield significant cross-linguistic difference, there was a noticeable tendency for each language group to choose more according to the language-specific semantics. More specifically, Korean speakers matched the target more often with loose (TT 5) or tight (TT 6) feature than did English speakers. (The cross-linguistic difference was more pronounced in detecting tight fit [TT 6] than in detecting loose fit [TT 5].) Conversely, English speakers matched the target more often with the containment feature than did Korean speakers. This may suggest that relative strengths at the perceptual/conceptual level for

detecting tight/loose and containment features differ between the two language groups and that the degree of strength is influenced by language-specific semantics (see more detail in Discussion).

Hypothesis 1b predicted that when containment and tight-fit compete as features for making similarity choices, English and Korean speakers would diverge, each categorizing more in line with language-specific semantics. As noted above, this hypothesis was tested in TT 8 and TT 9 where containment is pitted against tight-/looseness for categorization. In TT 8, the target relation is Loose-IN, and the choices are *tight-in* and *loose-on*. The *tight-in* choice shares the containment feature with the target, whereas the *loose-on* choice shares the looseness with the target. TT 8 indeed yielded a significant cross-linguistic difference: English speakers were significantly more likely to choose *tight-in* than were Korean speakers [Wald $\chi^2(1) = 11.34, p < .001$], matching the target in terms of containment rather than looseness. Korean speakers, in contrast, were significantly more likely to choose *loose-on* matching the target with looseness. The regression analysis for the Language contrast shows that about 7% of the variance in choices was associated with differences between the two language groups, which is about a medium size effect.

In TT 9, the target relation was Tight IN. In this TT, the *tight-on* choice shares tightness with the target, but the *loose-in* choice shares containment. This TT also yielded a significant cross-linguistic difference [Wald $\chi^2(1) = 22.25, p < .001$]: English speakers chose on the basis of containment more often than did Korean speakers, whereas Korean speakers chose on the basis of tightness (i.e., *tight-on*) more often than did English speakers. About 14% of the variance in choice scores was explained by differences between the two language groups. Interestingly, although English speakers matched the target with the containment feature more than Korean speakers did, they actually did not show a clear preference for containment over tight feature in their choices (average score of 2.90 for *loose-in* and correspondingly 3.10 for *tight-on*). This suggests that while perception of tight fit is strong (as shown in TTs 6 and 7), when the tight-fit feature competes with the containment feature for categorization, its salience gets much weaker for speakers of English, and they judge similarity based on the containment feature significantly more often than Korean speakers.

3.2. Support as target relation

Hypothesis 2 predicted that loose support will be perceived as distinct from tight support by both language groups (2a), but that the two groups will differ in the way they categorize tight support (2b): English speakers will pick out the support feature for categorization more often than will Korean speakers. The hypothesis was tested in TTs 10 through 18 where either loose support (TTs 10–13) or tight support (TTs 14–18) was the target relation with various types of contrast for the choices. Among these TTs, four of the five Triad types (TTs 15–18) involving tight support as the target relation yielded significant cross-linguistic difference (or closely approached significance). In contrast, all TTs with loose support as the target relation showed similar patterns of choice in the two languages. We now look at each part of the hypothesis, 2a and 2b, in more detail.

In Hypothesis 2a, we predicted that loose support would be perceived as distinct from tight support by both groups. This hypothesis was based on the infant studies showing that infants do not generalize a loose support relation to include a tight support relation (e.g., Choi & Casasola, 2005). The hypothesis was tested in TTs 10–12 and was supported. When Loose-ON was the target, both language groups were similar in the extent to which they viewed Loose-ON to be distinct from Tight-ON. In TT 10, where the choices were *loose-on* and *tight-on*, both groups chose *loose-on* to be more similar to the target than *tight-on*. Given that in English semantics, both choices are members of the category “on,” the preference for *loose-on* in the similarity judgment must come from the perceptual similarity among the members of the loose support category that sets it apart from those in the tight support category. Consistent with TT 10, *tight-on* was *not* the preferred choice in TT 11: English speakers considered *tight-on* to be equally different from Loose-ON as the other choice (i.e., *tight-in*). And in TT 12, both groups chose *loose-in* to be more similar to the target than *tight-on*.⁹ Perceptual distinction between loose-on and tight-on is also demonstrated in TT 14: When the target was Tight-ON and the choices were *tight-on* and *loose-on*, both groups chose *tight-on* to be more similar to the target than *loose-on*.

In TT 13, we tested whether loose support would be a category distinct from loose containment. As would be expected from the results of TT 4 that also contrasts the two relations (with loose containment as the target relation), loose support was indeed distinguished from loose containment.

In Hypothesis 2a, we also stated a possibility that English speakers perceive the two types of support—tight and loose support—to be more similar to each other than Korean speakers do. This possibility was based on the finding that language facilitates the development of a support category that includes both loose and tight support (Casasola, 2005b). The results of TTs 10–12, although not significant, show a consistent tendency in the direction of language-specific semantics: English speakers tended to match the target more with the support feature than Korean speakers did.

Hypothesis 2b predicted that categorization of tight support will be guided by language: English speakers will pick out the support feature for categorization more often than will Korean speakers, and Korean speakers will pick out the tight-fit feature more often than will English speakers. This hypothesis was tested in TTs 14–18. These TTs yielded significant cross-linguistic differences or closely approached significance (except for TT 14, which was explained in relation to Hypothesis 2a). Except for TT 14, about 6%–11% of the variance in choices was related to the language group difference, which is again a medium size effect. When the target relation was tight support, the two language groups differed basically across the board: Korean speakers made choices on the basis of tightness, whereas English speakers made their choices more on the basis of support. Let us look at each TT individually.

In TT 15, where the two choices were containment relations (*tight-in* vs. *loose-in*), Korean speakers were significantly more likely than English speakers to choose the scene that matched the target in terms of tightness (i.e., *tight-in*; Wald $\chi^2(1) = 11.42, p = .001$). In TT 16 where support and tightness were pitted against each other (*tight-in* vs. *loose-on*), Korean speakers were much more likely to choose tightness than English speakers and English speakers chose more on the basis of support than Korean speakers [Wald $\chi^2(1) = 8.97,$

$p = .003$]. Notice, however, that English speakers preferred to choose similarity by tight-fit feature (over support), but their preference was much less compared to Korean speakers. This suggests that while tight-fit is a strong perceptual cue for both groups, its activation for categorization (i.e., degree of strength) is much stronger for Korean speakers than for English speakers in relation to tight support. TT 17 (*tight-on* vs. *tight-in*) and TT 18 (*loose-on* vs. *loose-in*) contrasted containment versus support. For both, English speakers were significantly more likely than Korean speakers to choose the scene that matched the target scene in terms of support [Wald $\chi^2(1) = 12.80$, $p < .001$, TT 17; Wald $\chi^2(1) = 17.60$, $p < .001$, TT 18]. Interestingly, Korean speakers' choices in the two TTs (17 and 18) were at chance level, suggesting that Korean speakers consider *tight-in* and *tight-on*, *loose-in* and *loose-on* to be equally similar (TT 17) or equally different (TT 18) from the target. The cross-linguistic differences in TTs 15–18 taken together show that with Tight-ON as the target relation, Korean speakers indeed chose significantly more on the basis of tightness than did English speakers (TTs 15 and 16), and that English speakers made their choice on the basis of support significantly more than did Korean speakers (TTs 16, 17, and 18).

To summarize, the results of TTs 1–18 demonstrate the following: First, containment and tight fit within the containment relation are perceptually salient to both language groups, but language-specific semantics play a significant role when containment and tight/loose features compete for categorization. Second, while loose support is viewed as distinct from tight support by both language groups, categorization of tight support is guided by language-specific semantics: For tight support relations (e.g., tight encirclement, tight covering, tight attachment), English speakers pick out “support” as the unifying feature more than Korean speakers, whereas Korean speakers pick out “tight fit” as the unifying feature more than English speakers. In addition, there were marginal but consistent differences between the two language groups in the degree to which they detected the tight/loose feature within the containment relation (TTs 5 and 6) and the “support” feature when viewing a loose support event (TTs 10–12): Korean speakers tended to pick out the tight/loose feature more than English speakers, and English speakers tended to pick out the containment/support feature more than Korean speakers.

4. Discussion

The current study examined the relative contribution of perception/cognition versus language-specific semantics in nonverbal spatial categorization. In particular, it sought to sort out the spatial categories that are perceptually driven and those that are formed with guidance from language-specific semantics. It also sought to identify the context in which the two elements provide guidance in nonverbal categorization of spatial relation.

We studied monolingual speakers of English and Korean, two languages that differ significantly in the way they categorize spatial relations. English highlights a contrast between containment and noncontainment (support), whereas Korean highlights a contrast between tight fit and loose fit. In particular, the category of *kkita* (“interlock, put x into a tight-fitting, interlocking relationship with y”) in Korean crosscuts the categories of containment and

support in English such that it includes both tight containment and tight support. The semantic systems of the two languages provided us with four features to test: tight containment, loose containment, tight support, and loose support. We could, thus, investigate the issue of language and cognition in contexts where several features contrast against one another and where these features are differentially highlighted across languages.

We hypothesized that speakers would draw on *both* universal perceptual saliency and language-specific semantics in classifying spatial relations at the nonverbal level. Based on recent findings on the development of spatial cognition in infancy and early childhood, we constructed specific hypotheses for the four types of spatial relation: Containment and tight fit within the containment relation would be perceptually/conceptually salient features that all speakers will utilize for categorization. However, when the two features (containment vs. tight-fit) compete, language will guide the categorization. For the support relations, we predicted that while loose support would be perceptually distinguished from tight support by both language groups, categorization of tight support—a relation of diverse configuration—would be guided by language-specific semantics.

These hypotheses were tested in a triad format (a target video followed by two choice videos) where the four spatial relations were differentially contrasted in eighteen triad types. The participants' task was to choose the video that was more similar to the target they had just seen. Our analyses of their choices supported all of the above hypotheses. In the following, we discuss our findings in terms of the relative contribution of language and perception/cognition, and we do so separately for the containment and support domains.

4.1. Containment and tight/loose relations

In the containment domain, both language groups could categorize on the basis of the containment feature alone. Both groups could also detect degree of fit (tight or loose) within the containment relation. Salience of these features played a significant role when the two choices contrasted only by one of the features, containment (TTs 1–4) or tight fit (TTs 5–7). In these cases, both language groups chose on the basis of the contrasted feature, and language-specific semantics had little effect on their categorization.

Language-specific semantics contributed significantly when containment and degree of fit (tight or loose) were pitted against each other in the choices (TTs 8 and 9). That is, when one choice shared the containment feature but the other shared the tight/loose-fit feature with the target relation, the choices made by the two language groups differed significantly: English speakers based their choice on containment significantly more than did Korean speakers, and conversely Korean speakers based their choice on degree of fit (tight or loose) significantly more than English speakers. These TTs constituted medium effects of language (ranging from 7% to 14%).

There are two possible explanations for the significant role of language in these contexts. One is that it is the specific categorization context that prompts our mind to draw help from language. That is, while the perceptual salience for containment and tight fit is equally strong for both language groups, it is only in the context of competition that language intervenes and provides guidance. The other explanation, however, is that relative strengths (i.e.,

levels of activation) of the two features are different in the two language groups at the perceptual/conceptual level and that they have been influenced by the language-specific semantics through years of using the target language. More specifically, the perceptual ability to detect the tight/loose fit is relatively stronger in Korean speakers compared to English speakers. (The reverse would be the case for containment: The perceptual ability to detect the containment feature in a spatial relation is relatively stronger in English speakers compared to Korean speakers.) This may be the case considering the consistent tendency of cross-linguistic differences we found in the degree to which the two groups detected tight fit/loose fit or containment features even when there was no competition (i.e., TTs 5 and 6). The observed tendency is also consistent with the findings by Hespos and Piccin (2009) and McDonough et al. (2003) that adult Korean speakers are more sensitive to degree of fit than English speakers in a nonverbal task. (See also Hespos and Spelke [2004] who compared infants in English environments and adult English speakers and found that while the infants could distinguish between tight and loose containment English-speaking adults could not.) If degrees of strength of these features are different at the perceptual/conceptual level in the two language groups, then these differences would lead to different choices when the two features compete for categorization.

4.2. Support and tight/loose relations

In the support domain, the nature of interaction between perception and language was different from that of containment. Unlike the containment relations where the feature of “containment” (whether tight or loose) was salient, in the case of support relations, loose support was viewed to be distinct from tight support by both groups, suggesting that loose support forms a perceptual category of its own. This is particularly interesting in light of English semantics. In English, the spatial preposition “on” includes both loose and tight support, and infants attend to linguistic input from an early stage to form the category of support (Casasola, 2005b). The perceptual distinction that English speakers made between loose and tight support (TTs 10–12) along with similar results in Korean suggests that loose support has a high degree of perceptual coherence within its category (putting something loosely on a flat surface, cf. Fig. 4) that sets it apart from tight support.

In contrast to loose support, categorization involving tight support as target relation was significantly influenced by language-specific semantics. When the target relation was tight support (TTs 15–18), regardless of what was contrasted between the two choices, Korean speakers consistently chose more on the basis of tightness than did English speakers, whereas English speakers chose more on the basis of support. Again, these TTs constituted medium effects of language (ranging from 6% to 11%). These consistent and significant cross-linguistic differences suggest that when English speakers see tight support relations (e.g., tight encirclement, tight covering, tight attachment, cf. Fig. 3), they pick out “support” to be the unifying feature more than Korean speakers, whereas Korean speakers pick out “tight fit” to be the unifying feature more than English speakers. Considering the nonverbal nature of the task, we argue that

language-specific semantics has influenced the perceptual/conceptual organization of tight support in the two language groups.

Our findings are partially consistent with previous studies with adults (Hespos & Piccin, 2009; Hespos & Spelke, 2004; McDonough et al., 2003; Norbury et al., 2008). The present findings are consistent with Hespos and Piccin's (2009) finding to the extent that when adults are habituated to a tight covering relation, a type of tight support, and then were asked to judge its similarity to tight and loose containment, Korean speakers showed sensitivity to degree of fit while English speakers did not. In that study, however, the habituation for tight support was limited to one type of tight support (i.e., tight covering) and one exemplar of that type. Our present findings are also in line with Norbury et al.'s (2008) observation that when participants were familiarized with tight containment (Tight-IN), both language groups gave significantly higher rating for the *tight-in* test event than *loose-in* test event, showing that both groups were sensitive to the tight feature of the familiarized relation. However, in that study, no cross-linguistic differences were found: When the familiarization was Tight-ON (with an exemplar of tight encirclement), both groups gave significantly higher rating for the *tight-in* test event than *loose-in* test event. In our study, although both language groups showed sensitivity to the tight feature of Tight-ON, Korean speakers used the tight feature significantly more than did English speakers for classification (cf. TT 15). The difference between the present findings and those reported by Norbury et al. may be the result of the methodological issues discussed earlier. In particular, in Norbury et al.'s (2008) study, the task was a 10-point scale similarity rating, where there may be individual differences in rating criteria for expressing degree of similarity. Also while Norbury et al. restricted the choices to *tight-in* versus *loose-in*, and used the same (or the same type of) Figure and Ground objects across events, the triads in the present study contrasted different types of spatial relation and used a variety of objects in the stimuli. Thus, the present study surveyed a much broader array of distinctions and revealed more nuanced differences in how adult speakers perceive these events as similar.

4.3. Relation to studies on infant spatial categorization

The present study has shown a remarkable correspondence between patterns of development of spatial categories in infancy/early childhood and those of adult categorization. In particular, the spatial features (e.g., containment) that preverbal infants could detect for categorization were also salient to adult speakers in both groups, and those features that infants could detect with help from linguistic input (e.g., tight-fit feature in tight-fit support relation) were the features that showed cross-linguistic differences in adult speakers. In many ways, this is not surprising. As spatial understanding is a fundamental part of our cognition, much of it is developed during infancy and early childhood. Perceptual exploration for space starts from the beginning of life (see Baillargeon, 2002, for a review) and detecting spatial properties for abstract categorization begins during the first year of life (e.g., Casasola et al., 2003; Hespos & Spelke, 2004; McDonough et al., 2003). Infants also pay attention to linguistic input from the earliest

stage of language development to form some of the abstract categories. Because these categories are formed so early in life, the internal structure of these spatial categories should continue into adulthood in large part, and the present study confirms it. What the present study adds further are how perceptual features and language-specific semantics interact over time as speakers keep using the native language and to what extent language can influence degree of salience of those perceptual features.

4.4. Difference between containment and support

The current study found more cross-linguistic similarities in relation to containment than in relation to support (particularly tight support). Moreover, both groups viewed containment relation to be perceptually more coherent than support relation. More specifically, both language groups categorized containment as a single category, putting tight and loose containment together (TTs 1–4), but they did not do so for the support category as they viewed loose and tight support to be distinct.

This “decalage” between the two types of spatial relation suggests differences in the degree of diversity of the perceptual properties in the members of the two categories. As Choi (2006a,b) has pointed out, all containment events are perceptually consistent in that something goes into a Ground object that is concave (e.g., Bowerman, 2007). Thus, the saliency of containment is driven by perceptual uniformity (Levinson et al., 2003). Such salience could explain the considerable agreement among several languages about the membership of the semantic categories referring to containment: relatively small, moveable objects contained in a three-dimensional Ground object.

In contrast, tight support has a number of perceptually different configurations. As shown in Fig. 3 (see also the ON category in Fig. 2), tight support involves various types of support such as encirclement, attachment, covering, and clipping. Correspondingly, languages differ greatly in semantically classifying various types of “support” relations, and it is a challenging task to identify a clear typology or an implicational scale (similar to Berlin and Kay’s [1969] implicational scale for color terms) for these relations (Levinson et al., 2003). We argue that for these diverse relations, language plays a critical role in helping us organize them at the nonverbal level.

4.5. Relationship between language and cognition

Our results have shown that both universal cognition/perception and language-specific semantics guide our spatial categorization, and the resource that speakers rely on more for categorization depends on the type of spatial relation and the context in which different spatial features are contrasted.

As far as we know, this is the first study that has pinpointed the nature of interaction between perception/cognition and language-specific semantics in a nonverbal categorization of spatial relations in adult speakers. However, further studies are needed to confirm the present findings: In particular, further cross-linguistic studies on “tight support” with other languages (that categorize it differently from English and Korean) are needed to confirm whether lan-

guage indeed is a primary guide to the categorization of tight support for all speakers. Further studies with other languages may also identify other language-specific spatial features that play a role in spatial categorization. Research that further investigates the effects of experimental conditions on the interaction between Language and Triad Type is also warranted. While in the present study, the interactional pattern of language and perception/cognition was unaffected by condition, it would be interesting to examine how the nature of interaction changes when speakers actually engage in linguistic encoding during similarity judgment. In addition, more research is needed to pinpoint the extent to which language is used in nonverbal categorization tasks and the variables (e.g., degree of complexity of the task, experimental conditions) that influence degree of language use during such tasks. Finally, we also hope that future studies will uncover more on the dynamic interaction between cognition/perception and language-specific semantics in other domains besides space.

Does language influence cognition? Or is cognition independent of language? Recent studies in the spatial domain have generally approached these questions with a sense of necessity to provide a categorical “yes” or “no”: The Whorfian position has argued that language does influence nonverbal cognition (Levinson, 1996; Majid, Bowerman, Kita, Haun, & Levinson, 2004; Pederson et al., 1998), whereas the Modular view has argued that it does not (Li & Gleitman, 2002; Papafragou, Hulbert, & Trueswell, 2008; Papafragou et al., 2002). In contrast, our answer to *both* questions is “yes, but partially” in the case of spatial categorization. Universal perception/cognition and language-specific semantics both contribute to nonverbal spatial classification task in important and unique ways. While we use nonlinguistic perceptual/cognitive cues in categorizing spatial relations, we use our language in resolving problems that perception/cognition alone cannot. Language provides guidance for classification when cognitive/perceptual information is insufficient or ambiguous for categorizing spatial relations. In this way, language is an integral part of our cognitive domain of space.

Notes

1. Imai and Gentner (1997) investigated the extent of shape or material bias in learning novel words in English- and Japanese-learning children (as well as adult speakers of those languages). They found that both universal ontological knowledge and language-specific grammar played a role in children’s early word learning and that the uses of the two types of knowledge depended on the object type.
2. There are other morphological categories, for example, nouns and adverbs, where spatial relations are encoded, but they are more optional than the two primary morphological classes (Hickmann, Hendriks, & Champaud, 2008).
3. In addition, *nohta* “loose support” includes the meaning of “release x from hand.”
4. This was shown with a preferential looking paradigm: When English 29-month-olds were familiarized with Tight-IN, they could detect the tight-fit feature and thus preferred looking at tight-IN during the test trials, but when they were familiarized with loose-IN, they showed no preference between tight-IN and loose-IN test, suggesting that they could not detect loose-fit.

5. A total of 24 combinations are possible for triad types (4 target relations \times 6 possibilities for pairing up the four types as choices), but six were eliminated because of lack of cross-linguistic differences based on previous studies and semantic differences in the two languages. Six triad types that were possible but not included in the study are as follows: (1) Tight-IN [*tight-in* vs. *loose-on*], (2) Tight-ON [*tight-on* vs. *loose-in*], (3) Loose-ON [*tight-in* vs. *loose-on*], (4) Loose IN [*tight-in* vs. *loose-on*], (5) Loose-IN [*tight-on* vs. *loose-on*], and (6) Loose-ON [*tight-in* vs. *loose-in*]. Previous studies have shown (McDonough, Choi, Bowerman, & Mandler, 1998) that Tight-IN versus Loose-ON and Tight-ON versus Loose-IN were two maximally different categories that both English and Korean speakers discriminate in their respective language as well as in nonverbal sorting tasks. This accounts for the exclusion of four of the six triads (1–4 above). In the case of Loose-IN [*tight-on* vs. *loose-on*] and Loose-ON [*tight-in* vs. *loose-in*], the target is semantically differentiated from both alternates in both Korean and English: Korean has three different terms for these three categories, while English has one term for the target and one term for the two alternates. Thus, speakers of both languages may choose one or the other choice equally often but for different reasons.
6. English speakers sometimes used a verb alone (e.g., *stacking*) to describe a scene. Korean does not have prepositions, and the canonical word order in Korean is Subject–Object–Verb. Thus, in Korean, the object noun precedes the verb.
7. Gennari et al. (2002) also report that even in a “naming first” condition where the participants described the target videos (motion events juxtaposing Path and Manner of motion) prior to the experiment, little linguistic encoding occurred during nonverbal tasks, such as a recognition task. However, it is possible that the present study, which uses complex spatial events, may have triggered more linguistic encoding than previous studies on Motion events.
8. The R^2 values in Table 4 are based on a linear least squares analyses that regressed the mean choice scores for each Triad Type on Language.
9. The choice of *loose-in* for the target Loose-ON may be due to the way loose-in scenes were filmed. As the camera was positioned diagonally above the spatial event, the video showed quite clearly the Figure objects being put on the bottom surface of a container. The bottom of a container (e.g., plastic container of various sizes and shapes) was flat most of the time, and participants may have perceived objects to be loosely put on the surface (particularly in the context of the target relation Loose-ON). Nevertheless, the preference for choosing *loose-in* over *tight-on* suggests that participants considered Loose-ON to be very different from *tight-on* and that they chose on the basis of looseness rather than support.

Acknowledgments

We thank Laura Greenig, Beth Gravis, Jeong Ryu, and Chiwon Chun for their help with data collection. We also thank Marc Rohlfing for developing the computer program for the

similarity judgment experiment and Katharina Rohlfing for her contribution during the design phase of the experiment. Our thanks go to Jennie Pyers and Terry Regier for their very helpful comments on an earlier version of this paper. This study was supported by NSF BCS-0091493 and NICHD R03 HD043831-01 to the first author.

References

- Baillargeon, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 47–83). Malden, MA: Blackwell.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 60–79). Cambridge, MA: MIT Press.
- Bowerman, M. (2007). Containment, support, and beyond: Constructing topological spatial categories in first language acquisition. In M. Aurnague, M. Hickmann, & L. Vieu (Eds.), *The categorization of spatial entities in language and cognition* (pp. 177–203). Amsterdam, the Netherlands: John Benjamins.
- Bowerman, M., & Choi, S. (2001). Shaping meanings for languages: Universal and language specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge, MA: Cambridge University Press.
- Casasola, M. (2005a). When less is more: How infants learn to form an abstract categorical representation of support. *Child Development*, 76, 276–290.
- Casasola, M. (2005b). Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. *Developmental Psychology*, 41, 183–192.
- Casasola, M., & Bhagwat, J. (2007). Does a novel word facilitate 18-month-olds' categorization of a spatial relation? *Child Development*, 78, 1818–1829.
- Casasola, M., Bhagwat, J., & Burke, A. (2009). Learning to form a spatial category of tight-fit relations: How experience with a label can give a boost. *Developmental Psychology*, 45, 711–723.
- Casasola, M., & Cohen, L. (2002). Infant categorization of containment, support and tight-fit spatial relationships. *Developmental Science*, 5, 247–264.
- Casasola, M., Cohen, L. B., & Chiarello, E. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development*, 74, 679–693.
- Choi, S. (2006a). Influence of language-specific input on spatial cognition: Categories of containment. *First Language*, 26, 207–232.
- Choi, S. (2006b). Preverbal spatial cognition and language-specific input: Categories of containment and support. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Action meets word: How children learn verbs* (pp. 191–207). Oxford, England: Oxford University Press.
- Choi, S., & Casasola, M. (2005). *Preverbal categorization of support relations*. Presented at the Bi-annual Conference for the Society for Research in Child Development, Atlanta, Georgia.
- Choi, S., McDonough, L., Bowerman, M., & Mandler, J. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development*, 14, 241–268.
- Gennari, S., Sloman, S., Malt, B., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83, 49–79.
- Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Mapping the nature of change in spatial cognitive development*. Oxford, England: Oxford University Press.
- Hespos, S., & Baillargeon, R. (2001). Reasoning about containment events in very young infants. *Cognition*, 78, 207–245.

- Hespos, S., & Piccin, T. (2009). To generalize or not to generalize: Spatial categories are influenced by physical attributes and language. *Developmental Science*, 12, 88–95.
- Hespos, S., & Spelke, E. (2004). Conceptual precursors to language. *Nature*, 430, 453–456.
- Hickmann, M., Hendriks, H., & Champaud, C. (2008). Typological constraints on motion in English and French child language. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Özçalışkan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 209–224). Mahwah, NJ: Erlbaum.
- Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 67, 169–200.
- Levinson, S. C. (1996). Frames of reference and Molyneux's question: Cross-linguistic evidence. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 109–169). Cambridge, MA: MIT Press.
- Levinson, S. C., Meier, S., & the Language and Cognition Group. (2003). Natural concepts in the spatial topological domain – adpositional meaning in crosslinguistic perspective: An exercise in semantic typology. *Language*, 79, 485–516.
- Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, 83, 265–294.
- Lucy, J., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge, MA: Cambridge University Press.
- Lucy, J., & Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 465–492). Cambridge, MA: MIT Press.
- Majid, A., Bowerman, M., Kita, S., Haun, D., & Levinson, S. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8, 108–114.
- Malt, B., Sloman, S., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, 40, 230–262.
- McDonough, L., Choi, S., Bowerman, M., & Mandler, J. (1998). The use of preferential looking as a measure of semantic development. In C. Rovee-Collier, L. P. Lipsett, & H. Hayne (Eds.), *Advances in Infancy Research*, Vol. 12, 336–354. Norwood, NJ: Ablex.
- McDonough, L., Choi, S., & Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, 46, 229–259.
- Munnich, E., Landau, B., & Dosher, B. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition*, 81, 171–207.
- Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, 47, 121–148.
- Norbury, H. M., Waxman, S. R., & Song, H. (2008). Tight and loose are not created equal: An asymmetry underlying the representation of fit in English- and Korean-speakers. *Cognition*, 109, 316–325.
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108, 155–184.
- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, 84, 189–219.
- Pederson, E., Danziger, E., Levinson, S., Kita, S., Senft, G., & Wilkins, D. (1998). Semantic typology and spatial conceptualization. *Language*, 74, 557–589.
- Trueswell, J., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63, 64–82.

Appendix

Presentation of dynamic spatial relations on video

The design of the video event was consistent in all scenes. In the initial frame of the video, one Figure object (e.g., a Styrofoam clamp, cf. Fig. 5a) is already in the desired end state with respect to the Ground object (e.g., on a bar). The frame began with a hand holding another Figure object (i.e., another Styrofoam clamp) away from the Ground object. The video then showed the action of putting the Figure object into or onto the Ground object in a tight or loose relation. This action was repeated once more with a third Figure object. The video ended with three Figure objects being placed in the desired end state with respect to the Ground object(s).

Task instruction

The instruction in English was as follows:

You will be performing a similarity judgment task. In the center of the screen you will see a “target” movie followed by two “alternate” movies side by side. Your task is to choose which of the “alternate” is most similar to the target-movie you just saw.

(The same instruction was given to the Korean participants in Korean.)