



ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: [www.elsevier.com/locate/COGNIT](http://www.elsevier.com/locate/COGNIT)

## Who is crossing where? Infants' discrimination of figures and grounds in events

Tilbe Göksun<sup>a,b,\*</sup>, Kathy Hirsh-Pasek<sup>a</sup>, Roberta Michnick Golinkoff<sup>c</sup>, Mutsumi Imai<sup>d</sup>, Haruka Konishi<sup>c</sup>, Hiroyuki Okada<sup>e</sup>

<sup>a</sup> Department of Psychology, Temple University, United States

<sup>b</sup> Department of Neurology, School of Medicine, University of Pennsylvania, United States

<sup>c</sup> School of Education, University of Delaware, United States

<sup>d</sup> Department of Environmental Information, Keio University at Shonan-Fujisawa, Japan

<sup>e</sup> Department of Engineering, Tamagawa University, Japan

### ARTICLE INFO

#### Article history:

Received 6 December 2009

Revised 26 June 2011

Accepted 3 July 2011

Available online xxx

#### Keywords:

Event perception

Cross-language comparison

Prelinguistic constructs

Figure and ground

### ABSTRACT

To learn relational terms such as verbs and prepositions, children must first dissect and process dynamic event components. This paper investigates the way in which 8- to 14-month-old English-reared infants notice the event components, *figure* (i.e., the moving entity) and *ground* (i.e., stationary setting), in both dynamic (Experiment 1) and static representations of events (Experiment 2) for categorical *ground* distinctions expressed in Japanese, but not in English. We then compare both 14- and 19-month-old English- and Japanese-reared infants' processing of grounds to understand how language learning interacts with the conceptualization of these constructs (Experiment 3). Results suggest that (1) infants distinguish between figures and grounds in events; (2) they do so differently for static vs. dynamic displays; (3) early in the second year, children from diverse language environments form nonnative – perhaps universal – event categories; and (4) these event categories shift over time as children have more exposure to their native tongue.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Verbs and prepositions express relationships between the figures and grounds that unfold in events. Thus, when we talk of a skater (a figure) who *glides* across the ice (the ground), the specific action of *glides* entails a figure and a ground and is distinguishable from other potential interactions like *tripping*, *hopping* or *slipping*. The learning of relational terms like verbs is central to language acquisition because verbs form the fulcrum around which a sentence is constructed. Learning these words, however, is difficult because infants must not only parse events into

components like figures and grounds but also “package” these components in ways that are aligned with their native tongue. By way of example, English rarely conflates ground information within the verb (consider the verbs *swim* and *fly*) but so-called “ground verbs” in Japanese routinely encode the spatial configuration of the ground being traversed (e.g. *wataru* ‘go across’ implies that someone crosses a flat barrier dividing two points such as a road or a railroad track) (Muehleisen & Imai, 1997).

This research is at the intersection of event processing and language development. We ask how children process basic components of events at a time when they are at the cusp of word learning, and also when most children have amassed a native vocabulary of 50 or more words. When do infants demonstrate an ability to parse events into foundational components like figures (i.e., prominent agent undergoing the motion) and grounds (i.e., a reference point or a stationary setting) and how does exposure

\* Corresponding author at: Department of Neurology, University of Pennsylvania, 3400 Spruce Street 3 Gates, Philadelphia, PA 19104, United States.

E-mail addresses: [tilbe@mail.med.upenn.edu](mailto:tilbe@mail.med.upenn.edu), [tilbegoksun@gmail.com](mailto:tilbegoksun@gmail.com) (T. Göksun).

to their native language influence toddlers' interpretation of these event components?

### 1.1. Infants process components of events

During the first year, infants detect an object's motion (Haith, 1980), discriminate changes in patterns of motion (e.g., Bogartz, Shinsky, & Schilling, 2000), and use motion to parse actions in events (Baldwin, Baird, Saylor, & Clark, 2001; Hespos, Saylor, & Grossman, 2009; Sharon & Wynn, 1998; Spelke, Born, & Chu, 1983; Wynn, 1996). Once infants perceive the actions within events, they must also detect those aspects of events that are related to linguistic expressions (Clark, 2003).

Cognitive linguists (Jackendoff, 1983; Lakoff, 1987; Langacker, 1987; Talmy, 1985; Talmy, 2000) and developmental psychologists (e.g., Mandler, 1992; Mandler, 2004) have long proposed that a set of prelinguistic constructs is foundational to learning relational terms. A dynamic event is composed of semantic components such as *path* (the trajectory of the motion), *manner* (how an action is performed), *source* (the starting point), and *goal* (the endpoint) (Jackendoff, 1983; Talmy, 1985) that are labeled across the world's languages. Other foundational constructs refer to the spatial relations between objects such as *containment* (putting things in a container) and *support* (putting things on a surface) (Choi & Bowerman, 1991).

Research on the way that infants and adults process events for language is relatively recent (Casasola & Cohen, 2002; Choi, 2006; Golinkoff, 1981; Golinkoff & Hirsh-Pasek, 2008; Göksun, Hirsh-Pasek, & Golinkoff, 2010; Lakusta, Wagner, O'Hearn, & Landau, 2007; Malt & Wolff, 2010; Mandler, 1992; Mandler, 2004; Mandler, 2010; Parish, Pruden, Ma, Hirsh-Pasek, & Golinkoff, 2010; Pruden, Göksun, Roseberry, Hirsh-Pasek, & Golinkoff, 2011; Pruden, Hirsh-Pasek, Maguire, & Meyer, 2004; Pulverman, Golinkoff, Hirsh-Pasek, & Buresh, 2008; Shipley & Zacks, 2008; Wagner & Carey, 2005; Wagner & Lakusta, 2009). Thus far, the spatial and event components studied share a set of common characteristics (Göksun et al., 2010). First, they are all perceptually available (e.g., one can detect the manner of gliding by witnessing the event above) (Mandler, 2004). Second, because these components are represented in the world's languages, they seem to be universal (e.g., all languages seem to mark paths and manners in events; Jackendoff, 1983; Talmy, 1985). Third, and importantly, even though they are expressed universally, there are language-particular ways of encoding these semantic components in verbs and prepositions. For example, some languages tend to express path information with prepositions and manner information within verbs (e.g., *run out* in English), whereas others codify the same path of motion within the verb by using an optional manner in the adverb (e.g., *salir corriendo* 'exit running' in Spanish).

Linguistic descriptions of the semantic components of sentences are good starting points for uncovering what infants might know about dynamic events. The burgeoning literature in this area suggests that preverbal children notice and categorize spatial and event components such as path, manner, source, goal, containment, and support that are codified in verbs and prepositions (Casasola &

Cohen, 2002; Choi, 2006; Lakusta, Wagner, O'Hearn, & Landau, 2007; Pruden et al., 2004; Pulverman et al., 2008). Pulverman and her colleagues, for example, habituated English- and Spanish-reared infants (14- to 17-months) to an animated starfish performing both a *path* and a *manner* (e.g., a starfish twisting as it moves over a ball). Using a within subject design, infants saw four kinds of events: *control* (i.e., same event as habituation), *path* change (the starfish twisting under the ball), *manner* change (the starfish spinning over the ball), and *path and manner* changes (starfish flapping past the ball). Both English- and Spanish-reared infants increased their attention to path changes and manner changes relative to the control event, suggesting that they isolated these components within events (Pulverman et al., 2008).

These findings are intriguing given the differential "manner bias" in English. By way of example, English has been estimated to have many more manner verbs compared to Spanish or Turkish (Slobin, 2005). When describing short event clips (e.g., a boy crawling up a low hill or a girl jumping into a pool), English speakers produced 18 times more manner verbs than path verbs (Naigles, Eisenberg, Kako, Hightler, & McGraw, 1998). Results from a recent study suggest that before 3 years of age, English-, Spanish-, and Japanese-speaking children assume that a verb labels the path represented in the event (Maguire et al., 2010). By age 3 and beyond, they manifest language-specific patterns of verb construal such that English-speaking children are more likely to map a novel verb to the manner of the motion, compared to Spanish- and Japanese-speaking children. Perhaps infants initially and universally extract the same information from the events that they witness and later, once language is processed, attend differentially to the semantic components of events that are highlighted in their language.

Some support for this notion also comes from the study of containment and support relations. *Containment* is defined as the relationship that occurs when something is fully or partially surrounded by a container and it is captured by the English word *in*. *Support* refers to the contact of an object on a surface and is illustrated by the English word *on*. Using both looking times and reaching behavior as dependent variables, Baillargeon and her colleagues demonstrated that infants discriminate between events in ways that demonstrate an understanding of containment and support (e.g., Aguiar & Baillargeon, 1999; Baillargeon, Needham, & DeVos, 1992; Hespos & Baillargeon, 2001a; Hespos & Baillargeon, 2001b; Hespos & Baillargeon, 2006; Hespos & Baillargeon, 2008; Wang, Baillargeon, & Paterson, 2005).

If infants' event perception starts from a common base, we might expect that prelinguistic infants will be sensitive to these spatial distinctions – even if they are not lexicalized in their native language. This *concept-to-language hypothesis* proposes that event categories like path and manner or containment and support would be acquired before language has its influence (e.g., Göksun et al., 2010; Hespos & Spelke, 2007). As language meets dynamic events, it may dampen attention to some event components and highlight others. And indeed, this is what the literature suggests. The finding that English-reared babies

perceive containment and support events that are expressed in Korean lends credence to this perspective (Choi & Bowerman, 1991).

In a now classic study, Choi and Bowerman (1991) noted that Korean does not have English-equivalent words for *in* and *on*. Instead, it encodes containment and support depending on the *degree of fit* relation between two objects. For example, putting a ring *on* a finger and putting a book *in* a cover are described with the verb *kkita*, signifying the tight-fitting relationship between two objects. The contrasting verb *nehta* connotes loose-fit relations (i.e., put in, around or together loosely), such as putting a book *on* a table or putting a fruit *into* a bowl (Bowerman & Choi, 2001; Choi, 2006).

Hespos and Spelke (2004) asked whether English-reared 5-month-olds' might distinguish between these "Korean" tight- and loose-fit events in both containment and support relations even though they are not expressed in their ambient language. After familiarization (e.g., an object fits tightly into a container), the infants were presented with both a familiar relation (tight-fit) and a novel relation (loose-fit). Infants looked longer at the novel relation, suggesting a keen sensitivity to tight- and loose-fit relations. By the second year after birth, responses from English- and Korean-speaking children diverged in how they processed *kkita* ('tight-fit in') vs. *nehta* ('loose-fit in') spatial relations. Even though English-speaking children at the ages of 29 and 36 months decreased in sensitivity to the difference between tight- and loose-fit containment events, Korean-speaking children maintained those distinctions. Hence, language-specific aspects of these spatial categories influenced children's nonlinguistic sensitivity at least by 29 months of age. In addition, English-speaking 29-month-old children with more words in their vocabularies relative to their peers, or the ability to produce the word *in*, were less likely to perceive the difference in the Korean degree-of-fit relation as compared to low vocabulary children or to those who did not yet produce the word *in* (Choi, 2006). Thus, exposure to native language environment, coupled with children's knowledge of the specific prepositions that encode these relations, negatively correlate with the detection of non-native semantic distinctions.

This concept-to-language perspective can be contrasted with an extreme view that we might term the *language-to-concepts hypothesis*, suggesting that children are prompted to parse nonlinguistic events only as they learn language. This hypothesis favors Whorfian linguistic relativity (Whorf, 1956), which proposes that language itself influences the way people think. That is, language is a "tool" that enables children to find components in the events they witness; learning a language might help in constructing new concepts (e.g., Bowerman, 2007; Bowerman & Levinson, 2001; Choi & Bowerman, 1991). A recent paper by Gentner and Bowerman (2009) offers a middle ground approach proposing that some spatial categories might exist prelinguistically, with others that are less salient and are represented more rarely across languages demanding linguistic experience to be learned (see also Gentner, 1982).

Taken together, the literature suggests that prelinguistic infants notice and conceptualize spatial and event

components in ways that are conducive to learning all of the languages in the world regardless of their ambient language. This view parallels the research in infant discrimination of all of the sounds of language before they home in on the particular contrasts used in their native language (e.g., Eimas, Miller, & Jusczyk, 1987; Kuhl, 2004; Kuhl et al., 1997; Werker & Tees, 1984). That is, infants might discriminate and attend to a broad palette of event constructs that will later be refined and "semantically organized" with respect to the native language (for reviews see Göksun et al., 2010; Hespos & Spelke, 2007).

To date, however, only a small set of event components relevant to learning relational terms has been examined. This work expands that literature by probing three questions: (1) How do English-reared infants process figures and grounds in dynamic and static representations of events? (2) Are infants from different language environments similarly sensitive to aspects of events that are codified in languages around the world? and (3) Does sensitivity to aspects of events change as infants are exposed to their ambient language?

### 1.2. Figure and ground

The relationship between a static figure and a static ground was initially studied by Gestalt psychologists (e.g., Koffka, 1935; Wever, 1927) in terms of segmenting figures from grounds. Since then, the perception of figure and ground has been investigated in the literature on adults' event processing (e.g., Kimchi & Peterson, 2008; Peterson & Gibson, 1994). There is also research on infants' perception of figure and ground relations (e.g., Johnson & Aslin, 1998; Johnson & Mason, 2002; Kaufman-Hayoz, Kaufman, & Stucki, 1986). However, more studies are needed to assess infants' ability to differentiate between dynamic figures on various grounds and in real life settings as a prerequisite for learning relational terms.

Why is it important to study infants' processing of dynamic figures with respect to grounds? The discrimination of humans in dynamic events might be fundamental and associated with the concepts of agency and animacy that are constructed in the first 2 years of life (Eimas, 1994; Mandler, 1992; Mandler, 2004; for a review see Rakison & Poulin-Dubois, 2001). The interpretation of other people as agents (Johnson, 2000), for example, is related to understanding the role people play in causal events (e.g., Golinkoff, 1975; Golinkoff, 1981; Golinkoff & Kerr, 1978; Oakes, 1994; Poulin-Dubois & Shultz, 1990); the intentionality behind a person's movement (e.g., Phillips, Wellman, & Spelke, 2002; Woodward, 1998; Woodward, 2003) and the means agents use to attain goals (Sodian, Schoeppner, & Metz, 2004).

The *ground* in a dynamic event is a reference point in the form of an object or the stationary setting of the scene. Ground information is central for the linguistic encoding of motion events. In English, for example, the prepositions *over*, *into*, *through*, and *across* specify both a path that the *figure* follows and the spatial properties of the *ground* object. Hence, 'into' not only refers to a path that the figure moves along, but also indicates that the ground object is some kind of enclosure (Talmy, 2000). When the ground

is the setting where an action takes place, different relational terms implicitly encode different grounds. For example, ‘across’ implies a relatively stable surface that can be traversed, while ‘along’ implies a more or less horizontal principal axis (Jackendoff, 1992; Landau & Jackendoff, 1993).

Intriguingly, in some languages such as Korean or Japanese, *ground* information for stationary setting is specified within the verbs (Choi-Jonin & Sarda, 2007; Muehleisen & Imai, 1997). Japanese, for example, classifies motion path verbs into two categories: directional-path and ground-path verbs. *Directional path* (DP) verbs define the direction of motion relative to a starting point or goal (e.g., *hairu* ‘enter’, *iku* ‘go’, *kaeru* ‘return’, *kuru* ‘come’), and do not restrict the *ground* on which the motion occurs (Muehleisen & Imai, 1997). However, *ground-path* (GP) verbs such as *wataru* ‘go across’, *koeru* ‘go over’, and *nukeru* ‘pass through,’ incorporate properties of the ground along with the direction of motion (Beavers, 2008; Muehleisen & Imai, 1997; Tsujimura, 2006). The spatial geometry of the ground is the key to assigning the correct verb to a motion event. For example, in the sentence *Jun wa kawa/michi o watatta* ‘Jun crossed the river/street,’ the meaning of *wataru* ‘go across a flat barrier dividing two points’ implies that there is both a starting point and a goal, and that the ground is a flat extended surface such as a street (see Fig. 1a and 1b). The verb *wataru* could not be used to describe a person moving across grounds such as a field or a tennis court, grounds that have no clear borders or barriers that demarcate the two sides of the plane. Even though the tennis court is separated into sections, there is no clear barrier between its sections or around it. To describe a crossing action on these grounds, a more generic verb *tooru* ‘go across a continuous plane’ is used. Thus, even the same action that takes place against different kinds of grounds would demand that a different verb be used. Compared to DP verbs, these GP verbs are very specific in regards to the ground that they encode. However, both animate and

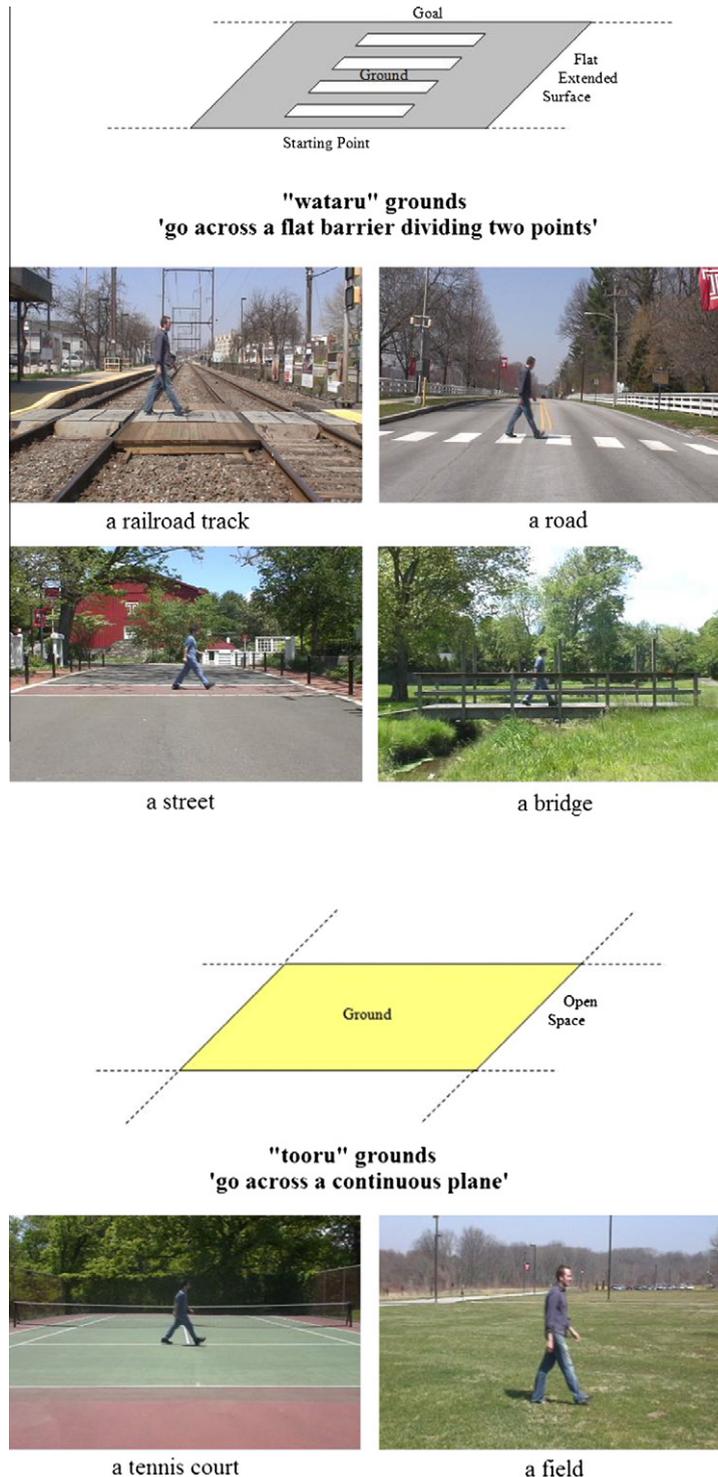
inanimate figures can be used (Muehleisen & Imai, 1997). In the current studies, we use only animate figures (i.e., humans) to show how crossing actions take place on different grounds. The use of all animate figures heightens the distinctions between the different kinds of crossing events without introducing another variable.

The encoding of grounds in some Japanese verbs, but not in English verbs, offers an opportunity to study another potential case in which event information is perceptually accessible, universally expressed in languages, and differentially labeled across languages. That is, as in the cases of containment and support or path and manner, the study of figures and grounds allows us to ask whether Japanese and English infants, prior to learning much language, make similar semantic figure and ground “cuts” when viewing dynamic events and whether exposure to language influences which ground distinctions are dampened or heightened when children learn their native tongue.

Using a nonlinguistic preferential looking paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hirsh-Pasek & Golinkoff, 1996), three experiments explore how English- and Japanese-reared infants discriminate figures and grounds in events and how native language exposure influences infants’ attention to these event components. In this paradigm, infants were first familiarized with single events and then shown one novel and one familiar event simultaneously. If infants can discriminate between figures and grounds in events, they should prefer to watch the *novel* figure or ground at test. Four specific questions were addressed: First, can English-reared infants detect figures in nonlinguistic dynamic events (Experiment 1)? We predict that infants will notice changes in the figures in dynamic events. Second, will English-reared infants show sensitivity to the detailed distinctions of grounds encoded only in Japanese (Experiment 1)? We predict that infants will be able to discriminate grounds with respect to the spatial geometry that is represented differentially within verbs in Japanese even though



Fig. 1a. Figures (a girl, a boy, a man, and a woman) used in the figure discrimination study.



**Fig. 1b.** Grounds used in the ground discrimination study. The grounds in the top panel (railroad track, street, road, and bridge) are encoded by the Japanese verb *wataru* 'a flat barrier dividing two points' and grounds in the bottom panel (tennis court and grass) are coded by the verb *tooru* 'a continuous plane'.

they are not coded in English. Third, is the detection of figure and ground different when infants are processing dynamic events vs. watching static scenes (Experiment 2)? We hypothesize that infants will differentiate between fig-

ures and grounds earlier when both components are static. Finally, how do English- and Japanese-reared toddlers diverge in their discrimination of grounds as they learn language (Experiment 3)? We expect that only Japanese-

reared toddlers who are exposed to a language that uses ground distinctions (*wataru* ‘flat barriers dividing two points’ vs. *tooru* ‘continuous plane’) will continue to notice the distinction between these classes of grounds.

## 2. Experiment 1: Discrimination of figures and grounds in dynamic events

### 2.1. Experiment 1a: Can 8- and 11-month-olds discriminate figures and grounds in dynamic events?

English-reared infants’ discrimination of figures and grounds in a *crossing* event was examined to ask whether infants notice language-relevant components in events even when their ambient language does not encode them. The Japanese ground-path verb *wataru* ‘go across a flat barrier dividing two points’ can be used with grounds such as a railroad track, a road, a bridge, and a street that extends in a line and with particular starting and ending points, but not with a tennis court or a grassy field that extend in a plane. Instead, the verb *tooru* ‘go across a continuous plane’ is used to describe a figure as it moves across a tennis court and a field. English, uses the verb ‘to cross’ appropriately with all six grounds. Thus, the Japanese verb *wataru* ‘a flat barrier dividing two points’ is more specific and restricted in meaning in comparison to the English verb ‘cross.’

Infants were familiarized with a dynamic scene in which a figure (e.g., a man) crossed a ground (e.g., a road). At test, they were presented with the same event in a split-screen, either with a change in the *figure* (e.g., a woman vs. a man crossing a road) or a change in the *ground* (e.g., a man crossing a railroad track vs. crossing a road). Importantly, ground comparisons were manipulated such that infants were presented with some grounds coded by the same verb in Japanese (e.g., a railroad track and a road) and with some examples that would be coded with two different verbs in Japanese: *wataru* ‘a flat barrier dividing two points’ as in the road and *tooru* ‘go across a continuous plane’ as in the tennis court. That is, infants were either presented with grounds that both extended in a line and were bounded (a railroad track vs. a road) or one that extended in a line and was bounded, and another that extended in a plane and was unbounded (a railroad track vs. a grassy field). Even though there are other grounds encoded by the verb *wataru* ‘go across’ (e.g., an ocean or a river, because these bodies of water also separate two points), for the current studies, we selected only a perceptually salient subset of grounds that all extend in a line (a road, a street, a bridge, and a railroad track).

Eight- and 11-month-old infants were tested in either a figure or a ground discrimination study. These age groups were chosen because previous findings suggest that infants discriminate various event components prior to producing their first words between 7 and 12 months of age (Lakusta et al., 2007; Pulverman et al., 2008). We hypothesized that (1) none of the infants in the figure or ground discrimination study would have an *a priori* preference before familiarization to a specific figure or ground; (2) infants should look longer to the *novel* figure or *novel* ground in the test trial, if they notice the change from familiarization trials

to test trial; and (3) parallel to the degree-of-fit distinction made in Korean, English-reared infants would also differentiate grounds better when the comparison was between two categories in Japanese (*wataru* ‘go across’ vs. *tooru* ‘go through’) as opposed to within the same Japanese category. That is, infants might consider the road vs. railroad track comparison more similar than the road vs. tennis court comparison.

### 2.2. Method

#### 2.2.1. Participants

Fifty 8-month-old ( $M = 8.01$ ,  $SD = .76$ , 24 males) and 45 11-month-old ( $M = 11.04$ ,  $SD = .82$ , 25 males) English-reared infants participated in this experiment. Infants were randomly assigned to either the figure discrimination (26 8-month-olds and 21 11-month-olds) or the ground discrimination study (24 8-month-olds and 24 11-month-olds). All infants were monolingual and full-term at birth. Infants were predominantly Caucasian and of middle-class families from two Northeastern cities in the United States. Infants who had been tested at 8-months-old were not tested again at 11 months. An additional 26 infants across two age groups were excluded from data analyses because they were bilingual ( $n = 7$ ), premature ( $n = 1$ ), the data were lost due to experimental error ( $n = 1$ ), low attention to video clips ( $n = 10$ ), having a side bias ( $n = 3$ , see below in the *coding* section for the criterion), or fussing-out during the experiment ( $n = 4$ ).

#### 2.2.2. Stimuli

The stimuli consisted of televised displays of four people (a woman, a man, a 6-year-old girl and a 6-year-old boy) crossing one of the six grounds (railroad track, road, narrow street, bridge, tennis court, and grassy field) from left to right. In a  $175 \times 125$  pixel image of the scene, children had an average height of 28 pixels and adults had an average height of 40 pixels. In each event, the figure crossed all the way across the ground two times, after one crossing was completed, as the clip repeated again from the beginning of the event. The pace of walking was controlled across the event clips. No linguistic audio accompanied these dynamic events. Stimuli were videotaped outdoors. All *figures* and *grounds* are presented in Fig. 1a and 1b.

In the figure discrimination study, different conditions displayed one of the three comparisons of figures: adult–adult, adult–child, and child–child. The figures were presented on different grounds in different conditions. For example, an infant might have seen the comparison of a woman and a man either both on a road or both on a tennis court. The use of human figures enabled us to keep the clips roughly comparable.

In the ground discrimination study, two conditions emerged based on the encoding in Japanese: within-category comparisons (*wataru* ‘a flat barrier dividing two points’ i.e., a railroad track, a road, a narrow street, and a bridge) and across-category comparisons (e.g., railroad track and tennis court or road and grass). The “*wataru*” grounds signal clear boundaries between the starting point and the goal point and extend in a line. A grassy field and a

tennis court are not proper grounds for the verb *wataru* 'a flat barrier dividing two points'; they are instead encoded by the verb *tooru* 'a continuous plane' in Japanese since they possess no barriers to divide them from their surroundings.

### 2.2.3. Procedure

Infants were tested using the nonlinguistic Preferential Looking Paradigm (Golinkoff et al., 1987; Hirsh-Pasek & Golinkoff, 1996; Pruden et al., 2004) where children are seated on their parents' laps approximately 2.5 feet from the front of a 44-in. television screen (34 × 28 in.). So as not to influence the child's direction of looking, parents were instructed to keep their eyes closed during the study.

Two cameras were placed at equal distance from the sides of the television (15 in.). The video camera on the right ran the movie for the study; the video camera on the left recorded the child's eye gaze for offline coding (for a similar set up, see Maguire et al., 2010). This enabled a clear view of eye gaze to the left or right side of the

screen. Once the experimenter started the movie, he or she left the room to avoid influencing the infant's attention. Each movie lasted for 132 s. Upon completion of the study, infants received a small gift of either a t-shirt or a toy.

The stimulus movie contained four main phases: introduction, salience, familiarization, and test trials (for sample conditions see Fig. 2). The trials were separated by an attention-getter.

### 2.2.4. Introduction phase

An animated character appeared first on one side of the screen and then on the other to ensure that infants were familiarized to clips playing on both sides of the screen. Each presentation was 6 s long.

### 2.2.5. Salience phase

Infants first saw what was to later become the test trial. This was used to determine whether there was any *a priori*

FIGURE DISCRIMINATION STUDY Visual Stimuli	
<i>Salience Trial</i> (12 sec)	 a man crossing a road  a woman crossing a road
<i>Familiarization</i> (4 times – 12 sec each)	 a man crossing a road
<i>Test Trial</i> (12 sec)	 a man crossing a road  a woman crossing a road
GROUND DISCRIMINATION STUDY Visual Stimuli	
<i>Salience Trial</i> (12 sec)	 a man crossing a road  a man crossing a railroad track
<i>Familiarization</i> (4 times – 12 sec each)	 a man crossing a road
<i>Test Trial</i> (12 sec)	 a man crossing a road  a man crossing a railroad track

Fig. 2. The design and sample stimuli for the figure (top panel) and ground (bottom panel) discrimination studies.

preference for either clip. The salience phase contained a 12-s clip of two events on a split-screen.

### 2.2.6. Familiarization phase

Infants watched four 12-s clips of exactly the same stimulus on the full screen that involved a figure crossing a ground (e.g., woman crossing a railroad).

### 2.2.7. Test phase

Infants watched the test events simultaneously on the split-screen for 12 s. In the figure discrimination study, infants were presented with the comparison between *same figure/same ground* (a woman crossing a railroad) vs. *novel figure/same ground* (a man crossing a railroad). In the ground discrimination study, the test trials compared the *same figure/same ground* (a woman crossing a railroad) clip seen during familiarization with the *same figure/novel ground* (a woman crossing a road) video clip. For within-category comparisons, the grounds were all from the “wataru” category, such as a railroad track vs. a road. For across-category conditions, the comparison used one ground from the “wataru” category and one ground that could not be described by the Japanese “wataru” (e.g., a road vs. a tennis court, or a railroad vs. a field).

### 2.2.8. Attention-Getter

A 3-s smiling baby face accompanied by the children’s song “Oh, Susanna” was used to separate the trials in each phase of the experiment. The attention-getter had two purposes: to renew infants’ interest in the movie and to reorient the infants’ looking to the center of the screen before they had to choose between pairs of stimuli in the split-screen trials. The side of the *novel figure* and the *novel ground* was counterbalanced in both salience and test trials. No linguistic stimuli or audio of any type accompanied the clips. Infants’ looking times were recorded for later coding.

### 2.2.9. Coding

The dependent variable was total looking time towards each event. During attention-getters and familiarization phases, infants’ looking to the center of the screen was measured. For the introduction, salience, and test phases, attention to the left or right side of the screen was coded using a button-press box (see McDonough, Choi, & Mandler, 2003).

Coders were always blind to condition to ensure that they did not know the target side in the movie. Each infant was coded twice to obtain intra-individual reliability. Coders were trained to consistently meet a standard of 99% reliability for both intra- and inter-judge coding. The intra-rater reliability is .998 ( $SD = .004$ ) for all participants; 41% of all videotapes were coded by a second person for inter-coder reliability ( $r = .991$ ,  $SD = .010$ ).

Attention was calculated by taking an infant’s total visual fixation time during all phases and dividing this number by the length of the entire movie. If infants looked at the movie less than 50% of the time (i.e., “low attention”), the data were removed from the sample. Ten infants were excluded due to this criterion. Side bias was calculated by dividing infants’ total looking to the right side of the screen

by their total looking time to both right and left sides of the screen. In this calculation, only split-screen phases were included. If the calculation was greater than .80 to one side, this was taken as an indication of a side bias, and the infant’s data would be excluded from analyses. Three infants were excluded due to this criterion.

## 2.3. Results

Percentage of looking time towards each event in the split-screen was calculated for *salience* and *test* trials. We report results separately for the figure and ground studies. As no significant gender differences emerged for either study, we did not consider gender as a separate factor for further analyses.

### 2.3.1. Figure discrimination

A repeated-measures ANOVA with age (8- and 11-month-olds) as the between-subject variable and trials (salience and test) as the within-subject variable yielded a trial by age group interaction,  $F(1, 45) = 7.14$ ,  $p = .01$ ,  $\eta^2 = .14$ . No main effects of trial or age group were found. That is, in the *salience* trials infants did not have any significant *a priori* preference for the event clips at any age: 8-month-olds:  $t(25) = 1.58$ ,  $p = .13$ , 11-month-olds:  $t(20) = .43$ ,  $p = .66$ . However, as seen in Fig. 3 (top graph), only 11-month-old infants looked significantly longer to the *novel figure* compared to the *same figure* in test trials,

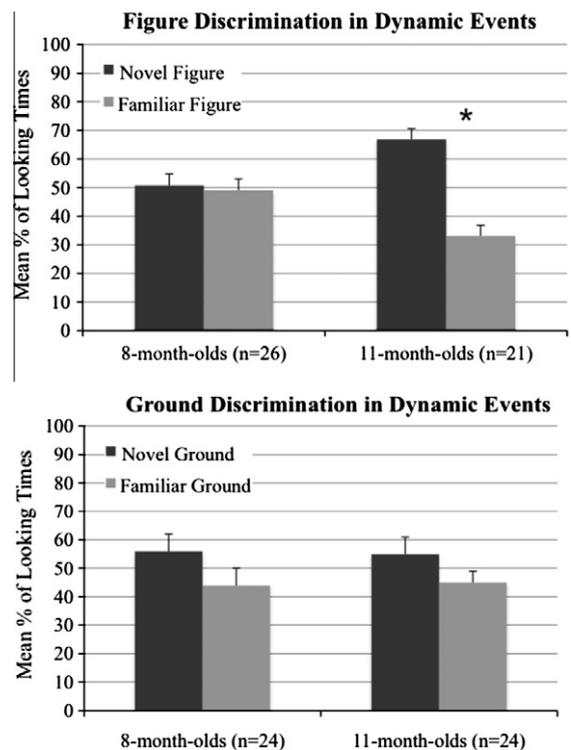


Fig. 3. Eight- and 11-month-old infants’ mean percentage of looking times in the test phase to novel vs. familiar figures (top panel) and novel vs. familiar grounds (bottom panel). \* $p < .05$ .

$t(20) = 4.53$ ,  $p = .01$  (two-tailed), which was also above chance level,  $t(20) = 4.20$ ,  $p = .001$ .

Infants in both age groups were equally attentive during the entire movie (71% and 74%, respectively for 8- and 11-month-olds). Additionally, looking times during the familiarization phase were examined to determine whether there was an age difference in infants' attention to the events. A repeated measures ANOVA with age (8- and 11-month-olds) as the between-subject variable and four familiarization trials as the within-subject variables yielded only a main effect of familiarization trial,  $F(1, 135) = 6.29$ ,  $p = .01$ ,  $\eta^2 = .12$ , but no main effect of either age or an age by trial interaction. Finally, no significant difference was found for different comparisons of figures (adult–adult, adult–child, or child–child).

### 2.3.2. Ground discrimination

A repeated-measures ANOVA with age (8- and 11-month-olds) and ground condition (within- and across-category) as the between-subject variables and trials (salience and test) as the within-subject variable demonstrated no main effects of trial, age group, or ground condition, nor any interactions among them,  $F_s < .114$ ,  $p_s > .23$ . Infants did not have any *a priori* preference for either event at either age: 8-month-olds:  $t(23) = .92$ ,  $p = .37$ , 11-month-olds:  $t(23) = .48$ ,  $p = .64$ , and infants at these age groups had similar percentages of looking times to *novel* and *familiar* grounds in the test trials in both within- and across-category comparisons (see Fig. 3, bottom graph). No significant difference was found for different comparisons of grounds in each condition (within- and across-categories; Fig. 4).

We again did not find any significant difference in terms of infants' attention during the entire movie (74% and 76%, respectively for 8- and 11-month-olds). Looking time analyses during the familiarization trials yielded no main effects of age or familiarization trial or an interaction between them. In both age groups, although there was a decline in looking times across the familiarization trials, it was not significant.

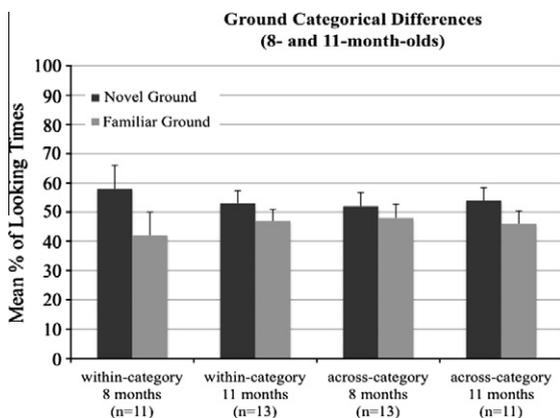


Fig. 4. Eight- and 11-month-olds' mean percentage of looking times to novel and familiar grounds at test in within- and across-category comparisons ( $p_s > .05$ ).

### 2.4. Discussion

The results from Experiment 1a suggest that by 11 months of age, English-reared infants only noticed the change of figures in these events. That is, infants discriminate figure changes earlier than ground changes in dynamic events. This is consistent with previous studies on other conceptual precursors such as containment-support and source-goal, which demonstrate that infants notice some components in events earlier than others. For example, infants distinguish the goal of an action before the source of an action (Lakusta et al., 2007) and containment events before support events (e.g., Casasola & Cohen, 2002).

Why do infants process figures earlier than grounds in dynamic displays? Figures are also expressed earlier than other conceptual distinctions in children's early utterance combinations (e.g., Clark, 1979; Grace & Suci, 1981; Tomasello & Merriman, 1995). Bock and Warren (1985) argue that the conceptual accessibility of what is related to a hierarchy of grammatical relations and the mental representations of these most accessible concepts are learned earlier. Research with toddlers has also demonstrated that visual attention was focused on the agents in a dynamic scene (Robertson & Suci, 1980). Hence, it is not surprising that in our study infants processed the human figures earlier than the grounds in the scenes.

On a related point, the perceptual saliency of event components might be related to differing developmental trajectories. Perhaps a moving, animate figure commands infants' attention more so than a stationary ground. Clark (2009) suggests that children's early conceptual categories are influenced by perceptual Gestalts. Thus, a figure is against a ground, and the moving figure object would be more salient than any one part of the scene. From very early on, infants are sensitive to motion and differentiate biological from nonbiological movement (Bertenthal, 1993). In addition, infants perceive the unity of a center-occluded object when it moves in front of a textured background surface, but they do not perceive object unity when it lacks a background texture (Johnson & Aslin, 1996; Johnson & Aslin, 1998; see also Kellman & Spelke, 1983). By the end of the first year of life, infants represent animate beings as having goal-directed movements, intentions, and as agents of causal events (e.g., Golinkoff & Kerr, 1978; Rakison & Poulin-Dubois, 2001; Woodward, 1998). The present study adds to this literature, suggesting that infants detect people as figures on various grounds in dynamic events.

Eleven-month-olds discriminated between all figure comparisons (adult–child, adult–adult, and child–child), suggesting that they detect information other than a person's height as a sign of figure change. Why were 8-month-olds not successful in recognizing the change of the person in these dynamic events? One possible explanation is that people were shown laterally, perhaps making it harder for younger infants to differentiate between them. Even though research suggested that 7-month-olds were more sensitive to actions than faces (Bahrick & Newell, 2008), in dynamic scenes where people's actions were similar (i.e., crossing different grounds), facial features might

be a useful and salient cue for discriminating between people.

### 2.5. Experiment 1b: Can 14-month-olds discriminate between grounds in dynamic events?

Experiment 1a suggested that 11-month-old infants noticed the change of a human actor in dynamic scenes on different grounds. In the ground discrimination study, however, we found that 11-month-olds did not have a significant preference for novel or familiar grounds at test. To examine the possibility of a developmental change in ground discrimination, we recruited an older age group of infants: 14-month-olds. We predicted that (1) none of the infants would have an *a priori* preference for specific grounds; (2) infants should look longer to the *novel* ground in the test trial, if they distinguish the ground change between the familiarization trials and the test trial; and (3) English-reared infants at this age would also differentiate grounds better when the comparison was *between* two categories in Japanese (*wataru* ‘go across’ vs. *tooru* ‘go through’). That is, these infants would notice ground changes better in across-category comparisons.

#### 2.5.1. Participants

Twenty-four 14-month-old ( $M = 14.22$ ,  $SD = .92$ , 12 males) English-reared infants participated. All infants were monolingual and full-term at birth. Infants were predominantly White and belonged to middle-class homes in two Northeastern towns in the United States. An additional six infants were excluded from data analyses due to experimenter error ( $n = 1$ ), infant low attention to video clips ( $n = 1$ ), or fussing-out during the experiment ( $n = 4$ ).

#### 2.5.2. Stimuli, procedure, and coding

All exactly the same as in the ground discrimination study in Experiment 1a.

### 2.6. Results and discussion

Percentage of looking time towards each event in the split-screen was calculated for *salience* and *test* trials. When no significant gender differences emerged, gender was not considered as a separate factor for further analyses.

A repeated measures ANOVA with ground condition (within- and across-Japanese category) as the between-subject variable and trials (salience and test) as the within-subject variable showed only a main effect of ground condition,  $F(1, 22) = 4.55$ ,  $p = .04$ ,  $\eta^2 = .17$ . No significant preference to either event was obtained at salience,  $t(23) = 1.17$ ,  $p = .19$ . However, infants looked significantly longer to the *novel* ground in the test trial,  $t(23) = 2.20$ ,  $p = .03$ . As shown in Fig. 5, infants looked significantly longer to the novel ground but only in the *across-category* comparison (e.g., railroad vs. tennis court) at above chance levels,  $t(12) = 4.48$ ,  $p = .001$ . They did not show the same pattern in the *within-category* comparison,  $p = .87$ . No significant difference was found for different comparisons of grounds within each condition.

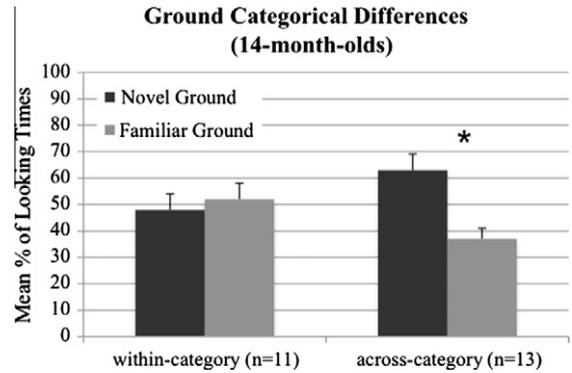


Fig. 5. Fourteen-month-olds' mean percentage of looking times to novel and familiar grounds at test in within- and across-category comparisons. \* $p < .05$ .

Lastly, looking times during the familiarization phase were examined to determine whether infants attended to the events and whether there was a difference in attention between within- and across-category comparisons. A repeated-measures ANOVA with condition (within- vs. across-category) as the between-subject variable and four familiarization trials as within-subject variables yielded only a main effect of familiarization trial,  $F(3, 66) = 7.59$ ,  $p = .01$ ,  $\eta_p^2 = .27$ . No main effect of condition or an interaction between familiarization trial and condition was found. This finding suggests that infants in both conditions were attentive during the familiarization phase and infants' looking times declined during familiarization across the trials.

Results from experiment 1b demonstrate that English-reared infants differentiated between grounds at 14 months of age. More importantly, when infants noticed ground changes, they did so for distinctions *not* encoded in English (i.e., the categorical difference of *wataru* ‘go across’ in Japanese). To what extent is this finding merely a product of the perceptual differences between these stimuli? If only perceptual differences mattered, then any variation in grounds should have led to looking time differences – even for comparisons of the grounds in the within-category conditions. This was not the case. Instead, these findings suggest that the geometry of grounds – despite perceptual differences – are better described as falling into distinct categories, namely, *wataru* and *tooru* categories.

Our findings corroborate the growing body of research on event perception. Infants parse events into their components and attend to distinctions in events even when these are not codified in their native language. Results from the ground discrimination study suggest that infants are sensitive to a categorical distinction between grounds made in Japanese but not in English. In particular, the grounds coded by the verb *wataru* ‘go across’ share certain geometric features: They are bounded and extend in a line with specific starting and ending points. In contrast, other grounds do not have these features. English-reared infants attend to these common features between *wataru* grounds, and notice the changes in grounds only when the ground changes from one Japanese category to another. This is reminiscent of English-reared 5-month-old infants'

sensitivity to the degree-of-fit relation encoded only in Korean (Hespos & Spelke, 2004).

### 2.7. Experiment 1c: Can 14-month-olds discriminate between grounds in dynamic events on grayscale?

The findings from Experiments 1a and 1b demonstrated that infants notice changes in figures at 11 months and grounds at 14 months in dynamic motion events. In ground discrimination, infants only noticed the *novel* ground when one ground was from the *wataru* category and the other was either a tennis court or a grassy field, grounds not encoded with the verb *wataru* 'go across.' Infants, for instance, differentiated a tennis court from a railroad track – a Japanese cross-category comparison – better than they did a railroad track vs. a road – a within-category Japanese comparison.

There is, however, a potential perceptual confound in the stimuli given the nature of the videos used to test this distinction. As shown in Fig. 1a and 1b, in addition to the nature of the geometry of the grounds, a tennis court and a grassy field have a salient color. One might argue that the homogenous stretches of green or red of the grassy field and the tennis court, respectively, allowed infants to categorize these two grounds together. To rule out this possibility and to secure our interpretation, we conducted a control experiment in which infants were presented exactly the same movies in grayscale. If color of the ground were a strong perceptual cue for infants' discrimination of grounds, we would not expect to obtain ground category effects using a black and white screen. Although removing the color will make the scene more artificial, findings from this control study would eliminate any possibility that infants look longer to grounds because of surface perceptual features.

We hypothesized (1) that there would be no difference between events in the salience trial, (2) if infants perceptually discriminate grounds according to the Japanese ground distinctions, removing color from the scene should not change the results. For example, infants would still find the comparison of road vs. grassy field (a cross-category comparison) easier to detect than the comparison of road vs. railroad track (a within-category comparison). Thus, we predicted that there would be no difference in ground differentiation between the color screen (Experiment 1b) and the black-white screen used here.

## 2.8. Method

### 2.8.1. Participants

Twenty-four 14-month-old ( $M = 14.01$ ,  $SD = .79$ ) English-reared infants participated, balanced for gender. All infants were monolingual and full-term at birth. Infants were predominantly White and from middle-class homes in two Northeastern towns. An additional 5 infants were excluded from data analyses due to bilingual exposure ( $n = 1$ ) and having low attention ( $n = 4$ ).

### 2.8.2. Stimuli

The stimuli were exactly the same as with the ground discrimination study in Experiment 1a, except that the

movies were presented in grayscale. Again, there were two conditions based on the Japanese GP distinctions: *within-category* (*wataru* 'go across', e.g., a railroad track and a road) and *across-category* (e.g., a railroad track and a tennis court or a road and a grassy field) comparisons.

### 2.8.3. Procedure and coding

The procedure and coding were the same as the ground discrimination study in the Experiment 1a. Twenty percent of all videotapes were coded by a second person for inter-coder reliability ( $r = .994$ ,  $SD = .005$ ).

## 2.9. Results and discussion

A repeated measures ANOVA with ground condition (within- and across-category) as the between-subject variable and trials (salience and test) as the within-subject variable yielded only a main effect of ground condition,  $F(1, 22) = 7.15$ ,  $p = .014$ ,  $\eta^2 = .25$ . No reliable preference for either event emerged in the salience trial,  $t(23) = 1.74$ ,  $p = .09$ , and no condition by salience interaction was found,  $p > .21$ . Only infants in the across-category condition preferred to look at the *novel* grounds at test,  $t(12) = 2.27$ ,  $p = .04$ , and this occurred at above chance level (see Fig. 6). No significant difference was found for different comparisons of grounds in each condition (within- and across-categories).

Another repeated-measures ANOVA with condition (within- vs. across-category) as the between-subject variable and 4 familiarization trials as within-subject variables yielded only a main effect of familiarization trial,  $F(3, 66) = 2.86$ ,  $p = .04$ ,  $\eta_p^2 = .12$ . No main effect of condition or an interaction between familiarization trial and condition were found. Infants in both conditions were equally attentive during the familiarization trials and their looking times declined across trials.

Thus, the results completely parallel those in Experiment 1b: when two grounds belonged to different categories as defined in Japanese (such as the road vs. the grassy field comparison), English-reared infants showed a preference for the *novel* ground on grayscale just as they had in the color exposure.

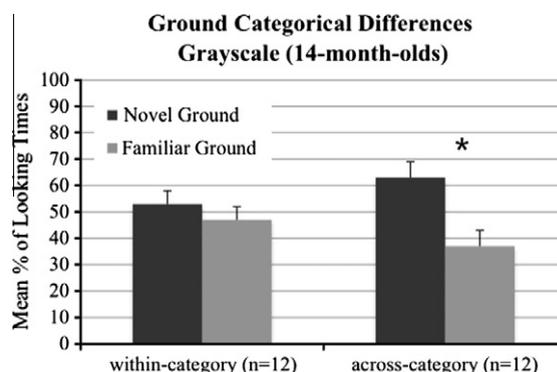


Fig. 6. Fourteen-month-olds' mean percentage of looking times to novel and familiar grounds at test in the within- and across-category comparisons in the ground discrimination study with grayscale. \* $p < .05$ .

We also compared the percentage of looking time to the novel ground between the infants who watched the movies on the color screen in Experiment 1b and in grayscale. Results showed that infants' looking time patterns were very similar between the two studies with no significant differences in looking time,  $p > .05$  (see Figs. 5 and 6).

The results of Experiment 1c argue against the possibility that infants were simply using the color of the grounds as a feature for ground categorization. Although there might be other perceptual confounds in the scene, the findings suggest that the *geometry* of the ground is a strong cue for differentiating between two grounds on a color screen and in grayscale. Hence, the nature of the ground, here defined abstractly as connecting specific starting and ending points and extending in a line or a plane, is noticed by English-reared infants when they process these dynamic events. Together with the findings from Experiment 1b, these findings suggest that the perceptual differences among grounds fall into two distinct categories, categories described by the Japanese verbs *wataru* and *tooru*.

### 3. Experiment 2: Discrimination of figures and grounds in static representations of events

Verb learning demands perceiving the spatial–temporal interaction inherent in dynamic events, because verbs capture a categorical moment in the unfolding event. For example, consider a woman running on the street. The woman runs in space, on a ground (i.e., the street) within a specific period of time. As she runs, both spatial and temporal dimensions change. Motion verbs are defined in large part by the interaction of figures and grounds across space and time. We thus wanted to explore whether infants perceive the semantic components of events in static displays of the same dynamic events. These scenes preserve the spatial dimension but not temporal aspect of those events. A picture of a dynamic event takes “a slice in time” as if the temporal dimension is *frozen* while keeping a static spatial configuration. Is the dynamic information important to the kinds of categories formed for verb learning or would static slices in time equally preserve the dimensions reflected in verb learning?

Past literature on containment, support, and degree of fit studies all utilized dynamic events showing a hand that placed objects into particular spatial relationships (Casasola & Cohen, 2002; Choi, 2006; Hespos & Spelke, 2004). Yet, the crucial distinctions between these events and ours are that in the past literature, discrimination of semantic elements could be made based on the *static end-points* of the events. For instance, “putting a toy tightly into a box” would be presented as an actual hand moving a toy tightly into a box, resulting in a static scene. In stark contrast, the ground-path verbs used in Japanese represent the interaction of the *path* of the *figure* against a particular type of *ground*. The verb *wataru*, for example, ‘flat barriers dividing two points’ from the dynamic ground discrimination study implies a dynamic motion on a particular ground rather than just a representation of the ground itself. It is possible that by teasing apart the spatial and temporal dimensions, static events might no longer represent

these scenes as different enough to be codified with different verbs (Muehleisen & Imai, 1997).

No research (of which we are aware) has investigated infants' ability to distinguish figures and grounds in static representations of dynamic events. Only one published study examined infants' attention to the relation between figure and context by using pictures of animals and vehicles on various grounds (Bornstein, Arterberry, & Mash, 2010). They found that 6-month-old infants categorized colored photographs of animals and vehicles across various contexts in which the figures appeared (e.g., a tiger in the green field, a tiger on a beach, and a tiger in a parking lot). The authors concluded that in category learning, infants initially focus on objects and ignore the contextual information.

Experiment 2 explored whether eliminating the temporal, but preserving the spatial aspect of an event enhanced or detracted from infants' ability to distinguish different figures and grounds. Instead of dynamic events, we used still shots of the people when they were in the middle of the screen. A static display might thus reduce attention to the figure and give priority to the ground, yielding an earlier discrimination of grounds. These studies allow us to specify whether infants notice the same semantic components to the same degree in static vs. dynamic events.

Eight- and 11-month-old infants' processing of figures and grounds was tested in static versions of the dynamic motion events used in Experiment 1a. These age groups were chosen to parallel those in the dynamic figure and ground discrimination studies (Experiment 1a). Also, given the findings of Bornstein et al. (2010), we expected that infants at these ages might be able to detect changes in figures and ground in static pictures. We expected that (1) there would be no *a priori* preference to static scenes before familiarization to a specific figure or ground; (2) similar to the dynamic experiments, in general, infants should look longer to the *novel* figure or the *novel* ground at test; (3) English-reared infants might not be sensitive to different ground distinctions (i.e., within- vs. across-category distinctions) in static events that no longer have a temporal dimension; and 4) infants might detect grounds earlier than they did in dynamic events, due to the presence of a static figure which might not attract as much attention as a dynamic figure.

#### 3.1. Method

##### 3.1.1. Participants

Thirty-five 8-month-old ( $M = 8.14$ ,  $SD = .85$ , 18 males) and 32 11-month-old ( $M = 11.12$ ,  $SD = .88$ , 13 males) English-reared infants participated. Infants were randomly assigned to the figure discrimination study (17 8-month-olds and 16 11-month-olds) or the ground discrimination study (18 8-month-olds and 16 11-month-olds). All infants were monolingual and full-term at birth. Infants were predominantly White and from middle-class homes in two Northeastern towns in the United States. An additional 22 infants across two age groups and figure and ground studies were excluded from data analyses because they were bilingual ( $n = 3$ ), premature ( $n = 1$ ), experimental

error ( $n = 4$ ), and having low attention to the video clips ( $n = 12$ ), or having a side bias ( $n = 2$ ).

### 3.1.2. Stimuli

This time, instead of dynamic events, we used screen shots of the figures (a woman, a man, a 6-year-old girl, and a 6-year-old boy) when they were in the middle of the screen (equally distant from both ends of the screen). All stimuli at all phases were static versions of the dynamic events. No linguistic audio accompanied these scenes.

As in the dynamic figure discrimination study, there were three comparisons of figures: adult–adult, adult–child, and child–child. The figures were also presented on different grounds within in each condition. For grounds, we again preserved two conditions based on how Japanese would encode these ground verbs in dynamic contexts. One condition involved the comparison of two grounds from the category of *wataru* (e.g., a road vs. a railroad track) and the other condition included cross-category comparisons with on ground from the “*wataru*” category and another from the “*tooru*” category (e.g., a road vs. a tennis court).

### 3.1.3. Procedure and coding

The procedure and coding for the static figure and ground discrimination studies were exactly the same as the dynamic figure and ground studies. Static stimuli were presented for the same amount of time as the dynamic stimuli. A second person coded 20% of all videotapes for inter-coder reliability ( $r = .994$ ,  $SD = .003$ ).

## 3.2. Results

We calculated infants' percentage of looking time towards each scene in the split-screen for *salience* and *test* trials. We report results separately for the figure and ground studies. No significant gender differences emerged for each study. Thus, gender was not considered as a separate factor for further analyses.

### 3.2.1. Figure discrimination

A repeated-measures ANOVA with age (8- and 11-month-olds) as the between-subject variable and trials (salience and test) as the within-subject variable yielded main effects of age and trial,  $F(1, 31) = 8.91$ ,  $p = .01$ ,  $\eta^2 = .22$  and  $F(1, 31) = 8.12$ ,  $p = .01$ ,  $\eta^2 = .20$ , but no significant interaction between them. Infants' looking times to the left and right side of the screen did not significantly differ in the salience trials at either age: 8-month-olds:  $t(16) = .99$ ,  $p = .34$ , 11-month-olds:  $t(15) = 1.06$ ,  $p = .31$ . As shown in Fig. 7 (top panel), only 11-month-old infants looked significantly longer to the *novel* figure compared to the *same* figure at test,  $t(15) = 3.03$ ,  $p = .01$ . Infants in both age groups were equally attentive to the whole movie (63% and 66%, respectively for 8- and 11-month-olds). Looking times during the familiarization phase yielded no main effects of familiarization trial, age, or an age by familiarization trial interaction ( $F_s < 1.98$ ,  $p_s > .08$ ).

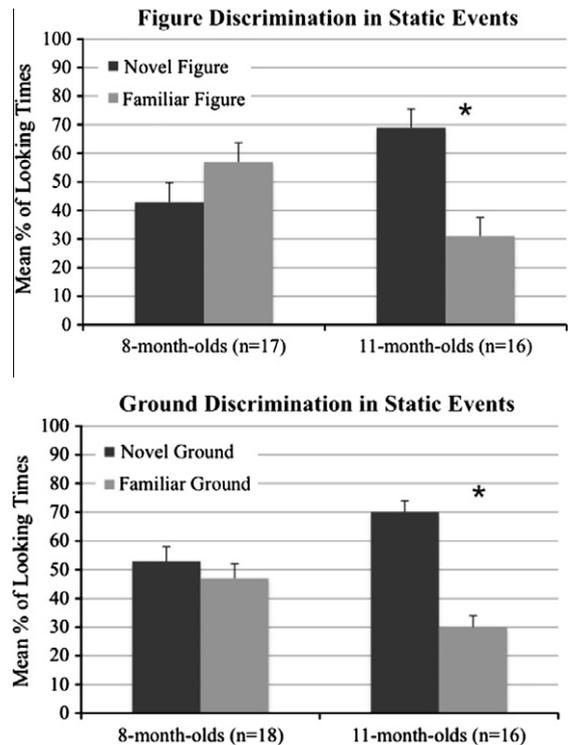


Fig. 7. Eight- and 11-month-olds' mean percentage of looking times in the test phase with static displays to figures (top panel) and grounds (bottom panel). \* $p < .05$ .

### 3.2.2. Ground discrimination

A repeated-measures ANOVA with age (8- and 11-month-olds) as the between-subject variable and trials (salience and test) as the within-subject variable yielded a main effect of trial,  $F(1, 32) = 8.38$ ,  $p = .01$ ,  $\eta^2 = .21$  and a trial by age interaction,  $F(1, 32) = 6.03$ ,  $p = .02$ ,  $\eta^2 = .16$ . That is, infants at both age groups had no significant preference for either event at any age in the salience trial: 8-month-olds:  $t(17) = .59$ ,  $p = .56$ , 11-month-olds:  $t(15) = .93$ ,  $p = .37$ . Only 11-month-old infants looked significantly longer to the *novel* ground compared to the *familiar* ground at test,  $t(15) = 5.45$ ,  $p = .01$  (see Fig. 7, bottom panel). Only the older age group (11-month-olds) appeared to notice the change in the ground in the test trial after familiarization, though the age groups were similar in their total attention during the whole movie (66% and 64%, respectively for 8- and 11-month-olds). Interestingly, and as predicted, infants detected the grounds in static scenes earlier than they noted ground changes in dynamic displays (Experiment 1b). Looking times during the familiarization phase yielded no main effects of familiarization trial, age, or an age by familiarization trial interaction ( $F_s < .84$ ,  $p_s > .44$ ).

Next, we analyzed whether infants were sensitive to the categorical ground distinctions coded by Japanese in static representations that lacked a temporal component. In stark contrast to what we found in Experiments 1b and 1c, infants started noticing the changes of grounds in static representations at 11 months of age and looked similarly to novel grounds *regardless* of the categorical distinction

in Japanese. Also, no significant difference was found for different comparisons of grounds in each condition,  $p > .05$  (within- and across-categories).

Finally, we tested infants' looking times during familiarization in the figure and ground studies for the dynamic vs. the static displays (the comparison of Experiments 1a and 2). Results indicated that infants looked significantly longer to the dynamic displays than to the static displays for both figures and grounds,  $F(1, 76) = 9.37$ ,  $p = .003$ ,  $\eta^2 = .11$ , and  $F(1, 78) = 13.96$ ,  $p = .001$ ,  $\eta^2 = .15$ . No significant interaction between age and study (dynamic vs. static) was found, suggesting that dynamic displays were more interesting to infants compared to static scenes.

### 3.3. Discussion

Experiment 2 demonstrated that English-reared infants differentiated static figures by 11 months of age, which was the same time they detected moving figures in dynamic motion events. In contrast, ground discrimination was detected *earlier* when the stimuli were static. In particular, infants noticed the change in grounds in static events 3 months before they did so for dynamic events. On the other hand, in the static versions of the dynamic events, infants did not distinguish between the types of grounds coded in Japanese. They treated *within-* and *across-category* comparisons of grounds similarly.

This experiment suggests that infants do not process static and dynamic events in the same way (Cutting & Proffitt, 1981). A critical finding, for example, was the disappearance of the categorical Japanese ground distinctions in the static versions of the same dynamic events. Without the dynamic motion, infants differentiated grounds both within- and across-categories, indicating that the perceptual distinctions between grounds in across-category comparisons were not more salient than within-category comparisons. But static scenes fail to capture the interaction between the figure and ground evident in dynamic scenes and therefore do not support the categorization of grounds. Arguably, the *movement* of the figure in dynamic events clarifies the starting and ending points of the grounds and whether the ground extends in a line or in a plane. The results also suggest that understanding verbs demands actual interaction between figures and grounds in actual events.

The removal of the dynamic aspects of the event also had the effect of enhancing infants' ability to distinguish between different grounds. Presumably, the person's movement in Experiment 1 oriented infants' attention to figures rather than grounds. Movement is a strong predictor of infants' attention to figures (e.g., Kellman & Spelke, 1983; Otsuka & Yamaguchi, 2003). Thus, younger infants appear to experience *change-blindness* as to "where" the action occurs when there are moving figures. Interestingly, during familiarization, infants had similar looking times in the figure discrimination and ground discrimination conditions for the static scenes, which were significantly lower than the looking times in the dynamic displays (which were also similar for the figure and ground condition).

Finally, this result raises questions about why infants successfully distinguish animals and vehicles in various

environments at 6 months of age in an earlier study (Bornstein et al., 2010), but detect static and dynamic human figures later. One reason that processing might be more difficult in this study in relation to Bornstein et al. is that we presented the humans using a lateral view of the figures on the screen, which made it more difficult to notice faces which might otherwise be more inherently interesting. Another explanation might be the size and location of the objects: the people were in the middle of the screen and the grounds occupied more space in comparison to the size of the figures in the static events.

### 4. Experiment 3: Cross-language comparisons on the detection of grounds

Our findings on infants' processing of grounds in dynamic events demonstrate that English-reared infants notice non-native ground distinctions. However, the evidence up to this point is inconclusive about the role language plays in infants' processing of these categorical ground distinctions. Would Japanese-reared infants process grounds in dynamic events in the same way as that evidenced by their English-reared peers? If so, it would offer support for the *concept-to-language* hypothesis. Testing Japanese infants also enables us to ask whether the early categorical distinctions infants notice in events change as they are exposed to their ambient language. Japanese children, immersed in a language that codes for these ground distinctions, should continue to notice different classes of grounds (e.g., *wataru* and *tooru*) as they become speakers of their language. English-reared infants, who do not hear these distinctions marked in their ambient language, should eventually dampen their ability to detect these event components.

This work parallels previous research demonstrating that children's sensitivity to non-native categorical distinctions decreases over time (e.g., Choi, 2006; Choi & Bowerman, 1991; Gentner & Bowerman, 2009). For example, after the second year of life, English- and Korean-speaking children responded differently to Korean tight- vs. loose-fit distinctions, suggesting that language-specific aspects of these spatial categories influence children's non-linguistic sensitivity. Thus, learning language dampens the detection of non-linguistic categorical differences that are not encoded in one's native tongue. In the final set of experiments we asked whether Japanese-reared infants differentiate between grounds in a way similar to English-reared infants and whether the early categorical distinctions infants notice in events change with exposure to their ambient language.

#### 4.1. Experiment 3a: Can 14-month-old Japanese-reared infants discriminate between grounds?

Here we examined how Japanese-reared infants processed grounds in dynamic events and whether they were receptive to the categorical ground distinctions encoded in Japanese. Fourteen-month-old Japanese-reared infants were tested because infants from English-speaking environments differentiate between grounds at this age. First,

it was hypothesized that there would be no preference for either event during the split-screen salience trials. Second, if infants discriminate between grounds in the test trial, they will look longer to the side where there is a *novel* ground. Last, similar to English-reared infants' responses, Japanese-reared infants would also differentiate grounds better when the comparison was between two categories in Japanese (*wataru* vs. *tooru*) as opposed to within the same Japanese verb category (*wataru*).

## 4.2. Method

### 4.2.1. Participants

The final sample was comprised of 26 14-month-old Japanese-reared infants ( $M = 14.07$ ,  $SD = .74$ , 15 males). Data were collected in the Greater-Tokyo Metropolitan area Japan by a trained Japanese–English bilingual experimenter. All infants were full-term and came from monolingual Japanese households. An additional five infants were excluded from data analyses because they fussed-out during the experiment ( $n = 4$ ) or had low attention to the video clips ( $n = 1$ ).

### 4.2.2. Stimuli, procedure, and coding

All were the same as those used in the dynamic ground discrimination study of Experiment 1. A second coding of 12% of subjects yielded high reliability ( $r = .992$ ,  $SD = .003$ ).

## 4.3. Results and discussion

A repeated measures ANOVA with ground condition (within- and across-category) as the between-subject variable and trials (salience and test) as the within-subject variable yielded only main effects of ground condition and trial,  $F(1, 24) = 16.32$ ,  $p = .001$ ,  $\eta^2 = .41$  and  $F(1, 24) = 7.37$ ,  $p = .012$ ,  $\eta^2 = .24$ , respectively. The interaction between trial and ground condition was marginally significant,  $F(1, 24) = 3.83$ ,  $p = .06$ ,  $\eta^2 = .14$ . No significant preference for either event emerged in the salience trial in either condition,  $t_s < 1.83$ ,  $p_s > .09$ . There was also no significant condition by salience trial interaction,  $p > .14$ . Only infants in the across-category condition preferred to look at the *novel* grounds at test,  $t(10) = 7.75$ ,  $p = .001$ , and this result was above chance level (see Fig. 8, bottom panel). Another repeated-measures ANOVA with condition (within- vs. across-category) as the between-subject variable and 4 familiarization trials as within-subject variables yielded only a main effect of familiarization trial,  $F(3, 72) = 4.44$ ,  $p = .006$ ,  $\eta_p^2 = .16$ . Neither a main effect of condition (within- vs. across-category) nor an interaction between trial and condition were found. Infants in both conditions were equally attentive during the familiarization trials and looking times declined across trials.

Finally, we compared English-reared infants from Experiment 1b and Japanese-reared infants in this experiment. Results indicated no main effect of language group or any interactions with language group,  $F_s < 1.09$ ,  $p_s > .30$  (see Fig. 8). Both age groups significantly differentiated between grounds only in the *across-category* conditions.

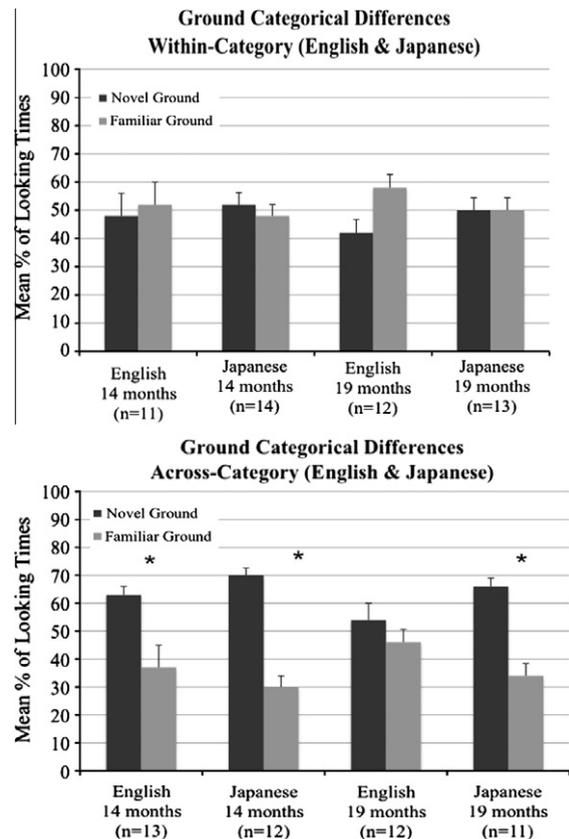


Fig. 8. English- and Japanese-reared 14- and 19-month-olds' mean percentage of looking times to novel and familiar grounds at test in within-category (top panel) and across-category (bottom panel) comparisons (\* $p < .05$ ).

The findings of this study suggest that 14-month-old Japanese infants and English-reared children of the same age demonstrated very similar sensitivity to distinctions between grounds in dynamic events. Just as English- and Korean-reared infants both detect the degree-of-fit relations in containment and support events even though this information was encoded only in Korean (Choi, 2006), English- and Japanese-reared 14-month-olds both attended to ground distinctions in nonlinguistic events in ways that are consistent with the category cuts used in Japanese.

These results support the *concept-to-language hypothesis* as well as Gentner and Bowerman's (2009) conjecture that some event categories might be acquired before language learning. An accurate assessment of these theories cannot be complete without testing the role of language learning on these event constructs. In the final experiment, we examine both English- and Japanese-reared older children's attention to the ground categorical distinctions in dynamic, nonlinguistic events.

### 4.4. Experiment 3b: Can 19-month-old english- and japanese-reared discriminate between grounds similarly?

Here we explore the link between learning a native language (i.e., English or Japanese) and the processing of

grounds in nonlinguistic events. If infants' attention to events alters as they learn their ambient language, we might expect differences between English- and Japanese-reared toddlers. Nineteen-month-old English- and Japanese-reared toddlers who generally pass the 50-word mark were recruited as participants and compared to those who were at the cusp of language learning in our prior experiments. At around 18 months of age, children seem to undergo a notable increase in the amount of vocabulary produced (e.g., Bloom, 1973; Dromi, 1987; Gopnik & Meltzoff, 1987; but see Bloom, 2000). Therefore, this is a good age at which to test the effect of language environment on processing nonlinguistic event categories.

We predict that Japanese children, immersed in a language that uses ground distinctions (*wataru* 'flat barriers dividing two points' vs. *tooru* 'continuous plane'), should maintain the distinction between these classes of grounds. As before, if toddlers discriminate between grounds in the test trial, they will look longer to the side where there is a novel ground.

#### 4.5. Method

##### 4.5.1. Participants

Twenty-four 19-month-old ( $M = 19.15$ ,  $SD = 1.02$ , 13 males) English-reared and 24 Japanese-reared toddlers ( $M = 19.28$ ,  $SD = 1.10$ , 11 males) from monolingual homes were tested. An additional 10 infants (17%) across both groups were excluded because they were bilingual ( $n = 1$ ), fussed-out ( $n = 7$ ), had low attention ( $n = 1$ ) or a side bias ( $n = 1$ ).

##### 4.5.2. Stimuli, procedure, and coding

All were exactly the same as the dynamic ground discrimination study of Experiments 1 and 3a. Inter-coder reliability on 12% of subjects was high ( $r = .995$ ,  $SD = .004$ ).

#### 4.6. Results and discussion

Results from the familiarization phase showed that both language groups and conditions (within- vs. across-category) were attentive. Looking times declined during familiarization across the trials,  $F(3, 132) = 8.48$ ,  $p = .01$ ,  $\eta_p^2 = .16$ .

A repeated measures ANOVA using language group and condition as the between-subject variables and two trials (salience and test) as the within-subject variables was calculated. There were main effects of language group and condition only on test trials,  $F(1, 44) = 4.34$ ,  $p = .04$ ,  $\eta_p^2 = .09$  and  $F(1, 44) = 8.42$ ,  $p = .01$ ,  $\eta_p^2 = .16$  (see Fig. 8) and no significant language group by condition interaction emerged. Neither English-reared nor Japanese-reared toddlers preferred either event in the salience trials,  $t_s < .84$ ,  $p_s > .41$ .

Planned pair-wise comparisons suggest that only Japanese-reared toddlers who were in the across-category condition looked significantly longer to the novel ground in the change trials,  $t(11) = 5.19$ ,  $p = .01$  (see Fig. 8, bottom panel). Japanese-reared children in the within-category condition and English-reared children in both conditions looked equally to both sides at test,  $t_s < 1.72$ ,  $p_s > .11$ .

Here, we predicted that only Japanese children, immersed in a language that supports ground distinctions, would maintain those distinctions whereas English-reared toddlers might dampen attention to contrasts not encoded in their native language. The results confirmed these predictions. Even though 14-month-olds from both English and Japanese language environments were equally sensitive to distinctions that the Japanese language makes among grounds in dynamic events (Experiments 1b and 3a), this effect was dampened for 19-month-old English-reared toddlers. These findings were not the consequence of general attentional differences between the language groups. One alternative possibility is that English 19-month-olds' lack of preference between grounds is a function of attentional biases toward the focal object/agent that have been documented in English-speaking cultures in adults (e.g., Masuda & Nisbett, 2001; Masuda & Nisbett, 2005). In other words, infants would not attend to ground features sufficiently to make the distinctions between ground categories. Future studies using eye-tracking methods will address this question. The current study offers preliminary evidence that language plays a role in the way in which we process events; if ground distinctions are not made linguistically (as in English) children's ability to detect these distinctions falters.

#### 4.7. General discussion

Figure and ground are foundational elements in events and are codified across languages. Whether you speak English or Turkish or Japanese, languages comment on events by frequently expressing the figure in the event and sometimes expressing the ground. Like other event components, figure and ground are perceptually accessible, universally recognized, and codified differently across languages (Golinkoff & Hirsh-Pasek, 2008). The current research expands the literature by studying the roots of these semantic distinctions in infant perception. In particular, we asked whether and how very young children perceive figures and grounds in events and how this perception might be modified when children start learning their native language.

Our findings indicated that, first, English-reared infants noticed changes in figures and grounds in dynamic events (Experiments 1a and 1b). Notably, infants were receptive to the non-native categorical ground distinctions for *crossing* action only in dynamic events. That is, the Japanese distinction proved to be a "default" for English-reared infants (Experiment 1b). Second, the geometry of the ground appeared to be a strong cue to distinguish between two grounds on both a color screen and in grayscale (Experiments 1b and 1c). Third, even though English-reared infants detected figures and grounds in static representations of the dynamic crossing events, the Japanese categorical ground differentiation no longer emerged in static displays, indicating that in these displays within- and across-category comparisons were treated similarly (Experiment 2). Finally, the early sensitivity to categorical ground distinctions by English- and Japanese-reared infants diverged as children began to process language patterns in their native languages. Japanese, but not English-reared toddlers preserved these distinctions, suggesting that the

process of learning language appears to shift early-formed categorical boundaries (Experiment 3). Together, the current experiments present evidence that infants parse non-linguistic dynamic events (and their static representations) into components to detect “*who is doing the action where.*”

Our findings on figure and ground detection in dynamic events add to the growing literature on infants’ perception of the foundational components of events (e.g., Casasola & Cohen, 2002; Lakusta et al., 2007; Pulverman et al., 2008), suggesting that by the latter half of the first year and the beginning of the second year, infants attend to nonlinguistic event components that are represented across languages. This is an essential first step in learning relational terms, particularly for motion verbs (Gentner & Bowerman, 2009; Golinkoff & Hirsh-Pasek, 2008; Göksun et al., 2010). Yet, some of these concepts seem to surface later in development than others (e.g., noticing figure changes before ground changes in dynamic events).

Once these conceptual foundations are present, they can be combined to create the semantic bases for word-to-world relations. For example, the early detection of the figures in dynamic events (e.g., Golinkoff & Kerr, 1978; Oakes, 1994; Robertson & Suci, 1980) allows children to name them early on (Clark, 1979; Grace & Suci, 1981; Tomasello & Merriman, 1995). One question is what truly constitutes a “figure” from an infant’s point of view. The presence of *people* in a scene might enhance infants’ attention, and thus their ability to notice changes. But will children have similar reactions to different figures such as cars, animals, and balls? Because the verb ‘crossing’ is permitted for all types of figures (e.g., the ball crossed, the person crossed, the dog crossed), children should respond similarly to different figure types in nonlinguistic events and extend their verbs to various figures. Future studies should tease apart the role of animacy in noticing the figures in different scenes.

Another finding was that changes in figures and grounds are differentially recognizable in static and dynamic displays. When the dynamic aspects of the events are removed, in general, infants’ ability to differentiate between grounds is improved. Nevertheless, the verb *wataru*, for example, ‘flat barriers dividing two points’ implies a dynamic motion on a particular ground rather than just a representation of the ground itself. Dynamic information with the attendant spatial and temporal interaction appears to be necessary to reflect “categories” for verb learning. A slice in time without the temporal component of an event is not enough to maintain categorical ground–path distinctions in events. In addition, preliminary results from a further study in which the path of the motion was modified from *crossing* to walking *alongside/near* the grounds, indicate that 14-month-old infants do not differentiate between grounds based on categorical distinctions when figures were walking *near* and not *across* grounds. This confirms the findings from the static and dynamic ground studies, suggesting that categorical ground distinctions occur only with the relevant interactions by a figure with ground and path (Göksun, 2010).

Even adults have difficulty detecting obvious changes in a picture of a scene when two pictures are presented sequentially (Simons & Levine, 1998). Nevertheless, cross-

cultural differences appear in the detection of changes in scenes (Ji, Peng, & Nisbett, 2000; Ji, Zhang, and Nisbett, 2004; Masuda & Nisbett, 2001; Masuda & Nisbett, 2005). Nisbett and colleagues investigated East Asians’ (e.g., Japanese, Chinese, and Taiwanese) and North Americans’ attention to the context and figures in a scene. East Asians were more attentive to relationships between objects and their environment than North Americans (Ji, Zhang, & Nisbett, 2004; Ji et al., 2000). Masuda and Nisbett (2001) also showed that compared to Americans, Japanese-speaking adults described the background of an underwater scene and expressed more relationships between the focal figure (e.g., a fish) and the background. In another study, using the *change blindness paradigm* (i.e., failure to detect the changes in a scene after exposure to it), Masuda and Nisbett (2005) displayed two animated scenes (e.g., a farm) that differed in small details. American adults again detected changes in the focal objects, but Japanese adults noticed the changes in the context and the relationships between the objects. Although these findings were mainly discussed in relation to cultural variations, the evidence corroborates our results on the difference between Japanese and American toddlers’ differentiation of grounds in dynamic events. Thus, these findings on change blindness could also be interpreted as a consequence of language rather than only cultural differences. Future studies are necessary to disambiguate these two interpretations.

#### 4.8. Conceptual distinctions before and after language

Despite differences in the ways in which languages encode foundational event components, infants seem to process these event constructs similarly before language has a chance to influence their perception (Hespos & Spelke, 2004; Hespos and Spelke, 2007; Göksun et al., 2010; Mandler, 2004; McDonough et al., 2003). The *ground discrimination* results with dynamic events support the claim that early event perception might be universal. English-reared infants are sensitive to the nature of the ground in a way that is more specific than encoded by the English verb ‘cross.’ This is consistent with findings from Hespos and Spelke (2004) that English-reared infants’ discriminate degree-of-fit relationships between two objects that are not encoded in English. In the current study, a conceptual distinction for grounds in dynamic events was revealed, demonstrating that infants noticed differences when a ground extends in a line (e.g., road) or extends in a plane (e.g., grassy field). That is, irrespective of the language environment in which infants are raised, they detect non-linguistic components of events, and infants attend to fine-grained distinctions in events even when these are not codified in their native language. The conceptual underpinnings for the learning of a language’s relational terms are in place as proposed by *concept-to-language* hypothesis (Gentner & Bowerman, 2009). As children note how these event components are lexicalized in their native tongue, they appear to tune into certain semantic distinctions over others, influenced by the ambient language. Future studies with English-speaking adults are planned to address the question of how readily these ground distinctions can be resurrected.

These findings are reminiscent of, though not analogous to, the universal phonological categories prelinguistic infants possess (e.g., Eimas et al., 1987; Kuhl et al., 1997; Werker & Tees, 1984). There might be a broad set of foundational components in events that will later be dampened by attending to only the subset that are coded in one's native language. While learning language, children might semantically reorganize their prelinguistic constructs by either subtly dividing their spatial and temporal world further or by creating a broader category (for details of this argument see Göksun et al., 2010; Hespos & Spelke, 2007).

A possibility that emerges from these findings carries implications for the learning of relational terms like verbs and prepositions in special populations. Perhaps problems in the learning of relational terms result partially from the inability to either perceive event components or reorganize event components with native language exposure. Children with autism, for example, might have particular problems with the processing of figures, because they show less attention to people in their environment. In contrast, children with autism might find some figures such as cars to be intrinsically more interesting than people (Sasson, Turner-Brown, Holtzclaw, Lam, & Bodfish, 2008). Thus, children with autism might respond differentially to the same event with different types of figures.

There is still much to explore on infants' ability to notice and abstract spatial and event components that are fundamental for learning verbs and prepositions. By bridging linguistics and event perception, we explored how infants process two components of events, figure and ground. We also selected perceptually salient ground-path categories. To generalize our conclusions, other grounds that are incorporated in verbs should be examined. More cross-linguistic studies – as well as studies with bilingual children – are required to confirm our assertions about the universal to language-sensitive shift. Studies across typologically varied languages about how children acquire the biases of their native language will illuminate the developmental links between language and thought.

## 5. Conclusions

Our findings begin to reconcile the long-standing debate on the role of language in shaping cognition, and provide support for concept-to-language hypothesis at least for very central and highly perceptual events constructs. That is, children's initial perception of events is not a "kaleidoscopic flux of impressions" as suggested by Whorf (1956) that awaits language for its organization. Rather, infants by 14 months are forming categories of the components of nonlinguistic events that appear to be the same regardless of the ambient language. Language appears to play a role in event perception once children notice how language 'underlines' different aspects of events.

## Acknowledgements

This work was supported by NICHD Grant 5R01HD050199 and by NSF grants BCS-0642529 to the

second and third authors, Japan Ministry of Education KAKENHI Grant 18300089 and Tamagawa University Global Center of Excellence (GCOE) program to the fourth and last author, respectively. Portions of this research have been presented at the 16th International Conference on Infant Studies, at the 11th Meeting of International Association for the Study of Child Language, and at the 33rd and 34th Meetings of Boston University Conference on Language Development. We would like to thank everyone at the Temple University Infant Lab, the University of Delaware Infant Language Project, Keio University Imai Lab, and Tamagawa University Baby Lab for their invaluable contributions to this project. Special thanks to Nora Newcombe for insightful comments on each study, Sarah Roseberry for discussions on many issues about the studies, Wendy Shallcross, Yannis Mitisziz, Katrina Ferrara, Russell Richie, and Aimee Stahl for their help in data collection. We would also like to express our deepest appreciation to all of the parents and infants who participated in the study.

## References

- Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology*, 39, 116–157.
- Bahrick, L. E., & Newell, L. C. (2008). Infant discrimination of faces in naturalistic events: Actions are more salient than faces. *Developmental Psychology*, 44, 983–996.
- Baillargeon, R., Needham, A., & DeVos, J. (1992). The development of young infants' intuitions about support. *Early Development and Parenting*, 1, 69–78.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, A. M. (2001). Infants parse dynamic action. *Child Development*, 72, 708–717.
- Beavers, J. (2008). On the nature of goal marking and event delimitation: Evidence from Japanese. *Journal of Linguistics*, 44, 283–316.
- Bertenthal, B. I. (1993). Infants' perception of biomechanical motions: Intrinsic image and knowledge-based constraints. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 175–214). Hillsdale, NJ: Erlbaum.
- Bloom, L. (1973). *One word at a time: The use of single-word utterances before syntax*. The Hague: Mouton.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Bock, K. J., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47–67.
- Bogartz, R. S., Shinsky, J. L., & Schilling, T. H. (2000). Object permanence in five-and-a-half-month-old infants. *Infancy*, 1, 403–428.
- Bornstein, M. H., Arterberry, M. E., & Mash, C. (2010). Infant object categorization transcends object-context relations. *Infant Behavior and Development*, 33, 7–15.
- 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: John Benjamins.
- Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–512). New York, NY: Cambridge University Press.
- Bowerman, M., & Levinson, S. C. (Eds.). (2001). *Language acquisition and conceptual development*. Cambridge: Cambridge University Press.
- Casasola, M., & Cohen, L. B. (2002). Infant categorization of containment, support, and tight-fit spatial relationships. *Developmental Science*, 5, 247–264.
- Choi, S. (2006). Influence of language-specific input on spatial cognition: Categories of containment. *First Language*, 26, 207–232.
- Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41, 83–121.
- Choi-Jonin, I., & Sarda, L. (2007). The expression of semantic components and the nature of ground entity in orientation motion verbs: A cross-linguistic account based on French and Korean. In M. Aurnague, M.

- Hickman, & L. Vieu (Eds.), *The categorization of spatial entities in language and cognition* (pp. 123–149). Amsterdam, Netherlands: Benjamin Publishers.
- Clark, E. V. (1979). Building a vocabulary: Words for objects, actions and relations. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 149–160). Cambridge: Cambridge University Press.
- Clark, E. V. (2003). *First language acquisition*. New York: Cambridge University Press.
- Clark, E. V. (2009). *First language acquisition* (2nd ed.). New York: Cambridge University Press.
- Cutting, J. E., & Proffitt, D. R. (1981). Gait perception as an example of how we may perceive events. In R. Walk & H. L. Pick (Eds.), *Intersensory perception and sensory integration* (pp. 249–273). New York: Plenum.
- Dromi, E. (1987). *Early lexical development*. Cambridge, England: Cambridge University Press.
- Eimas, P. D. (1994). Categorization in early infancy and the continuity of development. *Cognition*, 50, 83–93.
- Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In S. Harnard (Ed.), *Categorical perception: The groundwork of cognition* (pp. 161–198). Cambridge University Press.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Language, thought, and culture* (Vol. 2, pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwig, S. Özcaliskan, & K. Nakamura (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaacs Slobin* (pp. 465–480). New York, NJ: Lawrence Erlbaum Associates.
- Göksun, T. (2010). The 'who' and 'where' of events: Infants' processing of figures and grounds in nonlinguistic events. Unpublished dissertation, Temple University, Philadelphia, PA.
- Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). Trading spaces: Carving up events for learning language. *Perspectives on Psychological Science*, 5, 33–42.
- Golinkoff, R. M. (1975). Semantic development in infants: The concepts of agent and recipient. *Merrill-Palmer Quarterly*, 21, 181–193.
- Golinkoff, R. M. (1981). The case for semantic relations: Evidence from the verbal and nonverbal domains. *Journal of Child Language*, 8, 413–437.
- Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, 12, 397–403.
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14, 23–45.
- Golinkoff, R. M., & Kerr, J. L. (1978). Infant's perception of semantically defined action role changes in filmed events. *Merrill-Palmer Quarterly*, 24, 53–61.
- Gopnik, A., & Meltzoff, A. N. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, 58, 1523–1531.
- Grace, J., & Suci, G. J. (1981). The role of attentional priority of the agent in the acquisition of word reference. In Paper presented at the Society for Research on Child Development, Boston, MA.
- Haith, M. M. (1980). *Rules that babies look by: The organization of newborn visual activity*. Potomac, Maryland: Lawrence Erlbaum Associates.
- Hespos, S. J., & Baillargeon, R. (2001a). Reasoning about containment events in very young infants. *Cognition*, 78, 207–245.
- Hespos, S. J., & Baillargeon, R. (2001b). Infants' knowledge about occlusion and containment events: A surprising discrepancy. *Psychological Science*, 12, 141–147.
- Hespos, S. J., & Baillargeon, R. (2006). Decalage in infants' reasoning about occlusion and containment events: Converging evidence from action tasks. *Cognition*, 99, B31–B41.
- Hespos, S. J., & Baillargeon, R. (2008). Young infants' actions reveal their developing knowledge of support variables: Converging evidence for violation-of-expectation findings. *Cognition*, 107, 304–316.
- Hespos, S. J., Saylor, M., & Grossman, S. (2009). Infants' ability to parse continuous actions series. *Developmental Psychology*, 45, 575–585.
- Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, 430, 453–456.
- Hespos, S. J., & Spelke, E. S. (2007). Precursors to spatial language: The case of containment. In M. Aurnague, M. Hickman, & L. Vieu (Eds.), *The categorization of spatial entities in language and cognition* (pp. 233–245). Amsterdam, Netherlands: Benjamin Publishers.
- Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1983). *Semantics and cognition: Current studies in linguistics series, No. 8*. Cambridge, MA: The MIT Press.
- Jackendoff, R. (1992). *Languages of the mind*. Bradford: MIT Press.
- Ji, L., Peng, K., & Nisbett, R. E. (2000). Culture, control, and the relationships in the environment. *Journal of Personality and Social Psychology*, 78, 943–955.
- Ji, L., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology*, 87, 57–65.
- Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences*, 4, 22–28.
- Johnson, S. P., & Aslin, R. N. (1996). Perception of object unity in young infants: The roles of motion, depth, and orientation. *Cognitive Development*, 11, 161–180.
- Johnson, S. P., & Aslin, R. N. (1998). Young infants' perception of illusory contours in dynamic displays. *Perception*, 27, 341–353.
- Johnson, S. P., & Mason, U. (2002). Perception of illusory contours by 2-month-old infants. *Child Development*, 73, 22–34.
- Kaufman-Hayoz, R., Kaufman, F., & Stucki, M. (1986). Kinetic contours in infants' visual perception. *Child Development*, 57, 353–358.
- Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, 15, 483–524.
- Kimchi, R., & Peterson, M. A. (2008). Figure-ground segmentation can occur without attention. *Psychological Science*, 19, 660–668.
- Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt Brace Jovanovich.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., & Ryskina, V. L. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684–686.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakusta, L., Wagner, L., O'Hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language Learning and Development*, 3, 179–197.
- Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–238.
- Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.
- Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., Haryu, E., Imai, M., Vengas, S., et al. (2010). A developmental shift from similar to language-specific strategies in verb acquisition: A comparison of English, Spanish and Japanese. *Cognition*, 114, 299–319.
- Malt, B. C., & Wolff, P. (Eds.). (2010). *Words and the mind: How words encode human experience*. New York: Oxford University Press.
- Mandler, J. M. (1992). How to build a baby II: Conceptual primitives. *Psychological Review*, 99, 587–604.
- Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. New York: Oxford University Press.
- Mandler, J. M. (2010). The spatial foundations of the conceptual system. *Language and Cognition*, 2, 21–44.
- Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, 81, 922–934.
- Masuda, T., & Nisbett, R. E. (2005). Culture and change blindness. *Cognitive Science*, 30, 381–399.
- McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, 46, 229–259.
- Muehleisen, V., & Imai, M. (1997). Transitivity and the incorporation of ground information in Japanese path verbs. In K. Lee, E. Sweetser, & M. Verspoor (Eds.), *Lexical and syntactic constructions and the construction of meaning* (pp. 329–346). Amsterdam: John Benjamins.
- Naigles, L., Eisenberg, A., Kako, E., Hightner, M., & McGraw, N. (1998). Speaking of motion: Verb use in English and Spanish. *Language and Cognitive Processes*, 13, 521–549.
- Oakes, L. M. (1994). The development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, 30, 748–756.
- Otsuka, Y., & Yamaguchi, M. K. (2003). Infants' perception of illusory contours in static and moving figures. *Journal of Experimental Child Psychology*, 86, 244–251.
- Parish, J., Pruden, S. M., Ma, W., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). A world of relations: Relational words. In B. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 219–242). New York, NY: Oxford University Press.

- Peterson, M. A., & Gibson, B. S. (1994). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science*, 5, 253–259.
- Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, 85, 53–78.
- Poulin-Dubois, D., & Shultz, T. R. (1990). The infant's concept of agency: The distinction between social and nonsocial objects. *The Journal of Genetic Psychology*, 151, 77–90.
- Pruden, S. M., Göksun, T., Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2011). Find your manners: How do infants detect the invariant manner of motion in dynamic events? *Child Development*.
- Pruden, S. M., Hirsh-Pasek, K., Maguire, M. J., & Meyer, M. A. (2004). Foundations of verb learning: Infants form categories of path and manner in motion events. In A. Brugos, L. Micciulla, & C. E. Smith (Eds.), *Proceedings of the 28th annual Boston University conference on language development* (pp. 461–472). Somerville, MA: Cascadilla Press.
- Pulverman, R., Golinkoff, R. M., Hirsh-Pasek, K., & Buresh, J. S. (2008). Infants discriminate paths and manners in nonlinguistic dynamic events. *Cognition*, 108, 825–830.
- Rakison, D. H., & Poulin-Dubois, D. (2001). The developmental origin of the animate–inanimate distinction. *Psychological Bulletin*, 127, 209–228.
- Robertson, S. S., & Suci, G. J. (1980). Event perception by children in the early stages of language production. *Child Development*, 51, 89–96.
- Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S. L., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research*, 1, 31–42.
- Sharon, T., & Wynn, K. (1998). Individuation of actions from continuous motion. *Psychological Science*, 9, 357–362.
- Shipley, T. F., & Zacks, J. M. (Eds.). (2008). *Understanding events: From perception to action*. New York: Oxford University Press.
- Simons, D. J., & Levine, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review*, 5, 644–649.
- Slobin, D. I. (2005). Linguistic representations of motion events: What is signifier and what is signified? In C. Maeder, O. Fischer, & W. Herlofsky (Eds.), *Iconicity inside out: Iconicity in language and literature 4* (pp. 307–322). Amsterdam/Philadelphia: John Benjamins.
- Sodian, B., Schoeppner, B., & Metz, U. (2004). Do infants apply the principle of rational action to human agents? *Infant Behavior and Development*, 27, 31–41.
- Spelke, E. S., Born, W. S., & Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. *Perception*, 12, 719–732.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 57–149). New York, NY: Cambridge University Press.
- Talmy, L. (2000). *Toward a cognitive semantics. Concept structuring systems* (Vol. I, pp. i–viii, 1–565). Cambridge, MA: MIT Press.
- Tomasello, M., & Merriman, W. E. (1995). *Beyond names for things: Young children's acquisition of verbs*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tsujimura, N. (2006). *An introduction to Japanese linguistics*. NY: Blackwell.
- Wagner, L., & Carey, S. (2005). 12-Month-old infants represent probable endings of motion events. *Infancy*, 7, 73–83.
- Wagner, L., & Lakusta, L. (2009). Using language to navigate the infant mind. *Perspectives on Psychological Science*, 4, 177–184.
- Wang, S., Baillargeon, R., & Paterson, S. (2005). Detecting continuity violations in infancy: A new account and new evidence from covering and tube events. *Cognition*, 95, 129–173.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.
- Wever, E. G. (1927). Figure and ground in the visual perception of form. *The American Journal of Psychology*, 38, 194–226.
- Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1–34.
- Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, 6, 297–311.
- Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science*, 7, 164–169.