Developmental Psychology
2002, Vol. 38, No. 4, 604–614

0012-1649/02/$5.00 DOI: 10.1037//0012-1649.38.4.604

Young Children Can Extend Motion Verbs to Point-Light Displays

Roberta Michnick Golinkoff
University of Delaware

He Len Chung and Kathy Hirsh-Pasek
Temple University

Jing Liu
Women and Children's Hospital

Bennett I. Bertenthal
University of Chicago

Rebecca Brand
University of Oregon

Mandy J. Maguire and Elizabeth Hennon
Temple University

In the first study using point-light displays (lights corresponding to the joints of the human body) to examine children's understanding of verbs, 3-year-olds were tested to see if they could perceive familiar actions that corresponded to motion verbs (e.g., walking). Experiment 1 showed that children could extend familiar motion verbs (e.g., walking and dancing) to videotaped point-light actions shown in the intermodal preferential looking paradigm. Children watched the action that matched the requested verb significantly more than they watched the action that did not match the verb. In Experiment 2, the findings of Experiment 1 were validated by having children spontaneously produce verbs for these actions. The use of point-light displays may illuminate the factors that contribute to verb learning.

Running is running whether it is Carl Lewis circling a track or Grandma running to the telephone. To use a verb correctly, children must recognize an instance of an action when they see it, regardless of who is performing it or where it is taking place. The purpose of this article is to take a closer look at the processes children use to extend familiar motion verbs. Will young 3-year-olds, who already know many verbs, extend them to novel events created with the use of point-light displays? This study begins to probe the nature of children's verb representations by examining how children extend verb labels when the portrayal of the action is reduced to lights moving on a black screen. What remains in these point-light displays are light sequences based on the typical motion information associated with a given action. Just as in subjective contours, where a triangle seems to emerge when only its angles are visible (Kellman & Arterberry, 1998; Kellman & Shipley, 1991), point-light displays offer viewers apparent actors and objects without specifying their properties. Thus, children see no specific agent, no specific location, and no specific objects (e.g., no shovel in shoveling). What point-light displays *do* preserve are the semantic components of the verb such as *manner* (how an action is performed) and *path* (the trajectory an action follows; Talmy, 1985). We refer to the confluence of the semantic components that remain in point-light portrayals of action as the *verbal essence*. If children are able to extend verb labels across a category of action, they should be able to extend a familiar label to the actions portrayed in these point-light displays. Because children probably have never seen point-light displays, the use of these minimized stimuli provides a stringent test of whether children have abstracted the verbal essence of particular motion verbs.

Verbs are the architectural centerpiece of the sentence. Verb learning is of central importance for young language learners. Despite verbs' significance, however, the study of verbs has taken a back seat to the study of object nouns (Tomasello & Merriman, 1995). This neglect can be attributed to several factors. First, when English was the target language of much research, it appeared that acquisition of verbs developed later than acquisition of nouns (e.g., Bloom, Lightbown, & Hood, 1975). However, once other languages were studied (e.g., Chinese and Korean), it appeared that verbs were not universally acquired after object labels (e.g., Choi & Bowerman, 1991; Tardiff, 1996). Second, because languages partition the meanings of verbs in different ways, identifying universal features that account for verb learning across languages seemed a daunting task (Gentner, 1982; Gentner & Boroditsky, 2001). Finally, the field lacked methodologies needed to study children's perception of the complex, dynamic events that map verb labels to actions and relations. Only recently have researchers begun tackling these issues and turning their attention to the study of how children learn verbs and verb categories. Research on early

Roberta Michnick Golinkoff, School of Education, University of Delaware; He Len Chung, Kathy Hirsh-Pasek, Mandy J. Maguire, and Elizabeth Hennon, Department of Psychology, Temple University; Jing Liu, Women and Children's Hospital, Providence, Rhode Island; Bennett I. Bertenthal, Department of Psychology, University of Chicago; Rebecca Brand, Department of Psychology, University of Oregon.

This research was supported by National Science Foundation Grant SBR9601306 to Roberta Michnick Golinkoff and Kathy Hirsh-Pasek, by National Institute of Child Health and Human Development (NICHD) Grant HD 25455–07 to Kathy Hirsh-Pasek, and by NICHD Grant HD 16195 to Bennett I. Bertenthal.

We thank Michelle McKinney, Karen Arnold, Christopher Ramey, and the many undergraduate students who assisted us in the collection of these data. Michael Maratsos and Cindy Fisher gave us very helpful feedback on an earlier version of the article.

Correspondence concerning this article should be addressed to Roberta Michnick Golinkoff, School of Education, Willard Hall, University of Delaware, Newark, Delaware 19716–2922. E-mail: roberta@udel.edu

verb learning is still relatively sparse. To master the verb system in any language, children must conquer three preliminary tasks. First, they must attend to and individuate actions in their environment. Research suggests that infants are keenly aware of movement and use movement to individuate objects (e.g., Mandler, 1992, 1998) and actions (Sharon & Wynn, 1998; Wynn, 1996). Second, infants must be able to form categories of actions without language. The action of jumping, for example, refers to a decontextualized category of jumping motions that includes different kinds of jumps made by the same actor (e.g., Bob jumping off tables and jumping off chairs) and the same action performed by different actors (e.g., Bob jumping off the chair and Sally jumping off the chair). Third, children must be able to map words to actions and action categories. Consider how inefficient language would be if we needed a new label for each nonidentical instance of the action of jumping. The fact that infants have action words among their first words is testimony to the fact that they can successfully form word–action mappings (e.g., Bloom, 1993; Smith & Sachs, 1990). It is still unclear, however, what specific parameters guide the extension of action labels to action categories.

Within the field of lexical learning, explanations for verb acquisition emerge primarily from two theoretical positions, one emphasizing social-pragmatic views (e.g., Akhtar & Tomasello, 2000; Nelson, 1996) and another that focuses on children's use of constraints or biases (e.g., Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989). The former underscores how children learn to make word-to-world mappings by becoming apprentices to master word users in their environment. The latter stresses how children rely on word-learning principles to narrow down the possible meanings of words. With respect to nouns, for example, Golinkoff et al. (1994) posited a principle called *extendibility*: Words label not just the original exemplar but also a category of objects of like kind (e.g., Golinkoff, Shuff-Bailey, Olguin, & Ruan, 1995). For beginning word learners, "like kind" seems to be defined by the shape of an object but not its color (Hollich, 1999). Furthermore, as Hennon et al. (2000) have shown, rapid extension becomes the default position only at around 19 months of age. At 10 and 12 months, babies do not automatically extend a novel noun to other category members.

Golinkoff, Hirsh-Pasek, Mervis, Frawley, and Parillo (1995) argued that the principles of noun learning posited by Golinkoff et al. (1994) can, in somewhat different form, cover verb learning as well. For example, just as the principle of extendibility states that count nouns must label more than just the original exemplar, extendibility must work for verbs for the same reason: Much of the conceptual power of language is rooted in its capacity to categorize nonidentical events. Choi and Bowerman (1991) reported that children's early verbs seemed to rapidly extend beyond their original contexts of use for both English and Korean language learners. Unlike with nouns, however, we know less about the development of the principle of extendibility with verbs.

What is also unclear is the basis for young language learners' motion verb extensions. If children rely on seeing the same agent performing an action, they will not extend verbs appropriately. Furthermore, even if they do rely on seeing the same agent, they are unlikely to observe the action performed in the same way twice. Alternatively, if children rely on seeing the action performed in the same manner by different agents, they will also fail to extend correctly. "Running" is performed differently by Carl Lewis and Grandma.

More likely, children (and adults) extend verbs on the basis of an averaged representation of what that action looks like, its verbal essence. Marconi (1997) captured this idea: "Many verbs of motion have . . . *a typical appearance, a physiognomy*" (p. 159, italics added). This typical appearance may represent what Pinker (1989) labeled the "shape" of an action. But for motion verbs, the notion of shape seems to be composed jointly of four components, what linguists refer to as *figure*, *ground*, *manner*, and *path* (e.g., Frawley, 1992; Talmy, 1985). A motion event must at a minimum involve a figure (be it animate or inanimate) doing the motion against some ground, or location. Therefore, figure and ground do not distinguish between motion events. What does distinguish between motion events is the manner in which the motion is performed and whether the motion proceeds along a path relative to the ground. That is, *manner* captures *how* the motion is performed as well as its intensity and duration. Thus, the motion verbs "tapping," "knocking," and "pounding" are distinct in that they differ in intensity but not in path because they entail motion at a stationary location. Others, such as "walking" and "jumping," differ in manner (although both involve the legs) and path: Walking occurs along a forward path, whereas jumping takes place on a vertical path. That walking typically entails a forward path is signaled by the fact that speech about walking along another path requires the speaker to add a prepositional phrase, as in "walking in circles."

To extend a verb correctly to a newly observed instance of an action, children would have to abstract the verbal essence of an action. The untenable alternative would be that children might rely on seeing a specific agent do the action, on seeing the action performed in a particular manner, or perhaps on the involvement of a key object (as in "This must be 'knocking' because the surface is a door"). Reminiscent of Mandler's (1992, 1998) discussion of "image schemas" is the idea that children must lose the detail of the individual event to be able to recognize new instances and extend their verbs. For Mandler, events babies witness are "bleached" and stored as "image-schemas," or analog representations. These image-schemas, abstracted from children's interpretations of how objects move in space, help reduce infinitely varying perceptual displays to a limited number of meaningful schemas. If, for example, children include vertical motion in their image-schemas of the jumping action, this will help them organize their future encounters with variations of this action. It is important to note that although Mandler's theory accounts for the formation of action categories, it stops short of describing how labels get mapped onto or extended to these categories. Mandler's argument is in line with Gibson's (1966, 1979) view that in perceiving events we detect the invariants that persist from one event to another of the same type. Thus, although specific, situational features of an event may change across occasions (e.g., the agent, location, and objects involved), children may use their perception of what does not change across events to help them identify and classify actions of the same category.

The result of this process would yield an abstract and flexible verbal essence of the action to which a particular verb can be mapped. Thus, if children use the word "fall" to comment on their own fall and the fall of an object, they are indicating their recognition of "fall" as a general concept that captures a downward path of motion. Although the particular individual or object doing the falling is not a part of the verb's meaning, the downward path does represent the core, typical appearance of the action of "falling." Similarly, the word "run" refers to movement of the limbs (the manner of motion) on a forward path. When a child recognizes that "run" can be used to describe Fido and Daddy, the child has abstracted the invariants of running, namely, movement of the feet (whether two or four and in a particular way) on a forward path.

If children abstract the verbal essence of the actions that motion verbs name, perhaps they can interpret actions in minimalist point-light displays. Golinkoff, Hirsh-Pasek, et al. (1995) suggested that if children can extend action labels to actions depicted in point-light displays (Bertenthal, 1993; Bertenthal & Pinto, 1994; Soken & Pick, 1992), then they must possess abstract and flexible schemas for the verbs in question. Johansson (1973) first introduced point-light displays to study event perception and the perception of biomechanical motion. His classic study involved attaching small lights to the head and main joints of an individual's body and filming the person performing different actions (e.g., walking) against a dark background. In the absence of any featural information, such as clothing, facial features, or outlines of body contours, adult observers were able to recognize the moving dots as a person walking. Subsequent studies using these moving light displays with infants and adults demonstrated the significance of prior knowledge in the perception of biomechanical point-light images (Bertenthal, 1993). For example, Kozlowski and Cutting (1977) showed that adults could determine the gender of a point-light walker on the basis of past visual experience with the gait patterns of men and women. And Bertenthal (1993) showed that 3- and 5-month-olds could discriminate a canonical, point-light walker display from perturbed versions of such a display in which the moving point-lights were spatially or temporally scrambled. This finding is intriguing because low-level perceptual differences were equivalent in the canonical and perturbed displays.

If 3- and 5-month-olds can discriminate between canonical and noncanonical walkers in point-light displays, perhaps young children can actually discriminate between a pair of point-light displays depicting familiar verbs. Motion verbs (such as "jump" and "run") are acquired earlier than other types of verbs (Behrend, 1990), such as change-of-state verbs (e.g., "break"; Bloom et al., 1975), and they describe events in which children themselves have participated (Huttenlocher, Smiley, & Charney, 1983). Given the "bare" nature of point-light displays, this study provides a strong test of the hypothesis that children abstract the verbal essence of motion verbs.

Here we present two experiments in which we looked at children's ability to extend motion verb labels to point-light displays. In the first task, young 3-year-olds were presented with point-light images of eight familiar action verbs in the intermodal preferential looking paradigm (Golinkoff et al., 1987; Hirsh-Pasek & Golinkoff, 1996). In this paradigm, designed to test for language comprehension, children see two simultaneously presented video events. A linguistic stimulus that accompanies the events "matches," or describes, only one of the two events. The logic of this method is that children will watch the screen that matches the linguistic stimulus more than they will watch the screen that does not match—*if* and only if they comprehend the linguistic stimulus. Previous data collected with the use of this paradigm already have demonstrated that 17-month-old infants can discriminate between events depicting familiar motion verbs (e.g., pushing vs. bouncing) when performed by human actors (Golinkoff et al., 1987). Because children do not encounter point-light displays in everyday life, however, it was unclear whether they would be able to discriminate between verb pairs depicted in this format, and we therefore began this research with children around 3 years of age who already knew a number of verbs. If children can discriminate between events in point-light displays, we will be a step closer to uncovering what aspects of an action children abstract to attach labels to. We could then begin to examine what children include in the meaning of a verb such as "dancing" or "throwing."

Experiment 1

Method

The Intermodal Preferential Looking Paradigm

Each child sat on a parent's lap (usually the mother's) and watched a pair of videotapes balanced for salience (as judged from the adult perspective) on two side-by-side televisions separated by 30 in. (76 cm; see Figure 1). The videotapes were played in complete synchrony, with each of the trials on the two tapes beginning and ending at exactly the same time. At test, the tapes were accompanied by linguistic stimuli designed to "match" the display on only one of the screens. If children looked more to the screen that contained the named action than to the screen that contained the foil, we inferred that they were able to discriminate between the point-light verb displays.

Procedure

In the laboratory, children first played with an experimenter while the parent heard a description of the research and signed a human subjects release form. To avoid inadvertent parental influence, we instructed parents to remain silent for the duration of the videotapes and to wear cloth-covered glasses over their eyes during the actual testing. If they failed to comply, their children's data were discarded.

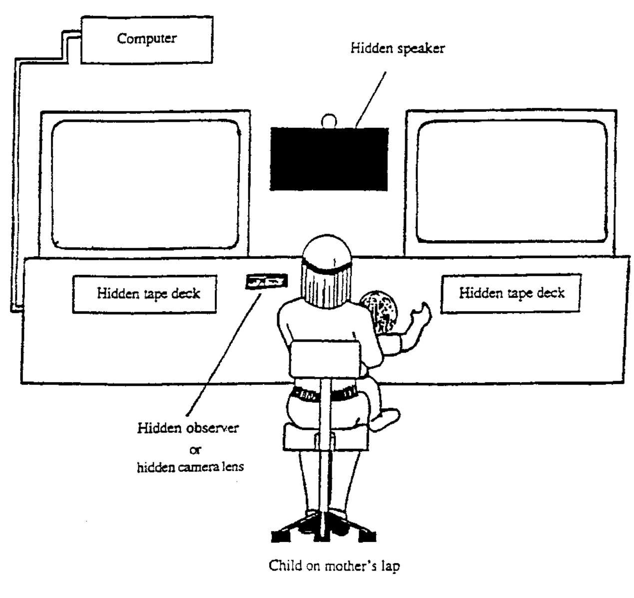

Figure 1. Diagram of the intermodal preferential looking paradigm.

The parent sat facing forward with uncrossed legs with his or her arms around the child; the parent was asked to keep the child centered on his or her lap and was reminded not to speak to the child during the video presentation. If the child got fussy, the parent was told to say something neutral to the child, such as "Just another minute," but to make no reference to the videos. The experimenter then darkened the laboratory, started the two videotapes, and left the testing area. Testing was discontinued if the child became fussy or upset. The parent was shown the tapes immediately after testing.

Participants

Although the majority of the sample was Caucasian, we contacted families on a race- and class-blind basis by sending letters to addresses obtained from the birth records available in local newspapers. There were 32 children (19 boys and 13 girls) between the ages of 35 and 38 months (range = 35.13 to 38.67 months; $M = 37.0$ months). To obtain the final sample, we tested 44 children. Data from 12 participants were excluded: 2 for fussiness, 2 for experimenter error or equipment malfunctions, 3 for side biases (greater than 75% visual fixation time to one screen side), 1 for insufficient looking (greater than 65% visual fixation to neither screen), 1 for too many failures to look at the center light between test trials (three or more), and 3 for parents failing to follow directions during testing.

Parents were first telephoned and asked whether their children understood the names of the eight actions appearing in the videotapes; production of the verbs was not required. To participate, children needed to comprehend seven of the eight verbs. If children did not meet this requirement, they were placed back into the pool of potential subjects and recontacted after a minimum of 4 weeks. Out of 32 children, 15 did not understand "skipping" and 3 did not understand "shoveling." The age selected for testing was determined empirically from these phone calls. Only at ages 35 months and older were we able to find children that knew seven of the eight verbs according to maternal report.

Structure of the Stimulus Videotapes

Biomechanical displays of motion verbs were created by videotaping a person with light-emitting diodes (LEDs) affixed to the major joints of the body (ankles, knees, hips, wrists, elbows, and shoulders). All actions were filmed in the dark and were devoid of all contour lines, human features, and objects (in the case of transitive verbs). These displays consisted only of a collection of white dots moving across a black screen. Figure 2 provides an example of a person walking, translated into frozen still images of point-lights. A cat video was also created by digitizing 90 frames of a videotape of a cat walking across a room, then importing the videotape into Adobe Photoshop, and placing white dots on specific anatomical locations. The brightness of the image was then reduced until only the white dots were visible.

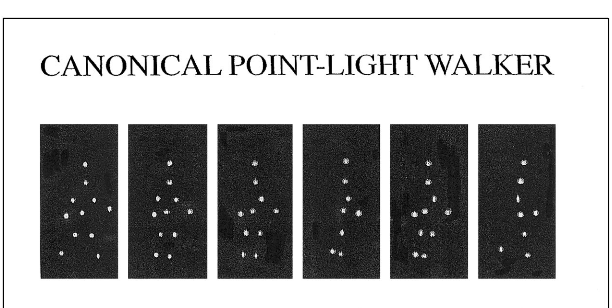

Figure 2. Illustration of a frame-by-frame (still) point-light display of a human figure walking.

Because no previous research had looked at whether children could recognize actions in point-light images, we included a range of motion verbs. The four pairs of images depicted walking versus dancing, shoveling versus picking flowers, running versus rolling, and hopping versus skipping (see Table 1 for more complete descriptions). Three of the four pairs consisted of intransitive verbs, and one pair was transitive. All point-light displays were created so that within a pair, the lights appearing in the displays were balanced for size, number, brightness, and movement.

To see whether children's ability to tell these displays apart was related to how similar they were, we asked adult college students ($n = 17$) to provide ratings of the perceptual similarities of these verb pairs when seen simultaneously, just as the children saw them. For the adult sample, however, no linguistic stimuli accompanied the tapes. Similarity within each pair was judged using a 7-point-scale. Thus, a score close to 7 meant that adults saw the actions in a pair as highly similar; a score close to 1 meant that the actions in a pair were considered much less similar. Table 4 shows these ratings.

Two separate tapes were created for the study. Tape 1 contained half of the actions (one from each pair of verbs), and Tape 2 contained point-light images of the other set of actions. Both videotapes in a pair were of identical length down to the number of frames, and tapes were synchronized to start playing at exactly the same time. Each pair of action verbs was presented in one of four separate blocks of trials that were identical in structure; one such block is shown in Table 2. Each trial in a block lasted 6 s and was separated by a 3-s intertrial interval during which the screens went blank. A female speaker recorded the linguistic stimuli in infant-directed speech for all trials, as well as between the trials, on one track of the videotape. The linguistic stimuli emanated from the center of the two monitors so that children could not use the television that played the stimulus as a cue to finding the matching screen.

Three types of trials. Before only the first block of trials, a brief, single *familiarization trial* was shown during which a point-light display of a cat walking across the screen appeared simultaneously on both monitors one time for 6 s. This trial served two purposes. First, it introduced children to the point-light stimuli before they needed to respond to test questions about them. Because point-light displays are not ordinarily encountered by children, it was important to offer them an opportunity to see these images before asking them to watch a particular action during test trials. Second, because it was not clear whether the children would be able to interpret these displays, we labeled the object and action appearing on the screens (i.e., "See the cat walking?"). We did not, however, familiarize children with stimuli that closely resembled the test images.

In order to ease children's transition into interpreting the contents of point-light displays involving human actors, the first pair of video events following this familiarization trial included the verb "walking" paired with the verb "dancing." Unlike the cat video, which filled the entire screen and moved horizontally from right to left, the human point-light walker was smaller in size (as were the other actors) and crossed the screen diagonally from left to right (see Table 1).

The familiarization trial was followed immediately by two *salience trials* during which different actions appeared simultaneously on each screen, one action on the left monitor and its mate on the right. Salience trials had three purposes: First, they showed children that contrasting events could appear on both screens at the same time. Second, they were used to calculate stimulus salience. If the actions in a pair were truly balanced for salience, visual fixation would be distributed approximately equally to both screens during this trial. Finally, the linguistic stimulus in the salience trials provided exposure to the names of the actions without telling children which screen either action was on, as in, for example, "Hey, one is walking and one is dancing!" Thus, children were directed to watch both screens. To maintain children's interest, different carrier phrases (e.g., "Whoa! See that?" or "Look at them!") were used throughout the tape to deliver the linguistic information.

Table 1
Descriptions of the Four Pairs of Actions Depicted in the Point-Light Displays

Videotape 1	Videotape 2
Walking: Person walking across the screen diagonally from the top left corner to the bottom right	Dancing: Person twisting in place
Shoveling: Person bending at the waist and standing back up (without moving the legs) as if shoveling snow and making a pile on his side	Picking flowers: Person bending at the waist and standing back up (without moving the legs) as if picking flowers up from the ground and then putting them in a hand-held basket
Rolling: Person performing one somersault across the screen diagonally from the top left corner to the bottom right	Running: Person running across the screen diagonally from the top left corner to the bottom right
Skipping: Person skipping across the screen diagonally from the top left corner to the bottom right	Hopping: Person hopping on one foot across the screen diagonally from the top left corner to the bottom right

Note. All displays repeated the actions continuously during the 6-s trials; the point-light figures that traversed the screen diagonally did so two times during any one trial.

The two *test trials* that followed the salience trials were designed to determine if children could distinguish between the displays and watch the action that matched the linguistic stimulus. Now children were exhorted to watch the screen containing the named action, as in "Look at dancing! See dancing?" The target action appeared on the same screen side for both test trials in a block, the same side as in the two preceding salience trials.

Intertrial intervals. Each trial was separated by a 3-s intertrial interval during which both screens went blank. A red light mounted centrally between the two televisions lit up during this time to attract the children's attention to the center, off the screens. This procedure ensured that children would not just remain fixated on one screen for long periods of time but would have to choose which screen to look at for each trial. The appropriate linguistic stimulus for the trial to follow was first heard during the intertrial interval so that prior to each test trial, the child was directed to find the matching screen (e.g., "Can you find dancing?").

Apparatus and Data Coding

All equipment—except for the two 19-in. (48-cm) color monitors—was shielded from the child's view. The videotapes were shown on VCRs. A 1-kHz tone was recorded for the duration of each trial on the second, inaudible channel of the videotape and was "read" by a specially designed tone decoder that functioned in two ways: (a) It turned the centrally mounted red light on during the intertrial intervals and off during the trials, and (b) it signaled the beginning and end of each trial to the computer (either an Apple IIe or an IBM 486).

Coders, blind to the sides on which the matching and nonmatching displays appeared, determined children's visual fixation to the two screens off-line from a silent videotape. Equipped with a button in each hand corresponding to the left or right video screen, coders depressed a button for the length of time a child looked at that screen. A look to the center was indicated by depressing both buttons; a look away from the screens was indicated by depressing neither button.

Dependent and Counterbalanced Variables

The dependent variable was mean visual fixation time (calculated to tenths of a second) to the named action (the match) versus to the foil (the

Table 2
Order of Events and Linguistic Stimuli for a Single Block of Trials

Videotape 1	Linguistic stimulus	Videotape 2
	Familiarization trial	
	[Center light]	
Blank screen	"Hey kids! Look at the cat!"	Blank screen
Cat walking	"Do you see the cat walking? Look at the cat!"	Blank screen
	Salience trials	
	[Center light]	
Blank screen	"Now you'll see some men!"	Blank screen
Walking	"One is walking and one is dancing!"	Dancing
	[Center light]	
Blank screen	"Did you see that?"	Blank screen
Walking	"They're walking and dancing!"	Dancing
	Test trials	
	[Center light]	
Blank screen	"Can you find dancing?"	Blank screen
Walking	"Watch dancing! Where's dancing?"	Dancing
	[Center light]	
Blank screen	"Do you see dancing?"	Blank screen
Walking	"Find dancing! Look at dancing!"	Dancing

Note. All three trial types (familiarization, salience, and test) were 6 s in duration. The center light was illuminated during each 3-s intertrial interval at the same time that the screens went blank. The light drew the child's attention to the center, off the screens.

nonmatch) during each pair of the test trials. Visual fixation times to the match and the nonmatch were averaged for each pair of test trials. For each test trial (test trials came in pairs of two), visual fixation time was measured, starting during the intertrial interval, from the point at which a child watched the center light for 0.3 s or more. Coding began during the intertrial interval because (a) by that time children had been told the names of both actions and (b) the audio exhorted the child to find the match during that interval. If the child failed to reach the 0.3-s intertrial interval criterion on a trial, that trial was not included in the data analysis (this occurred on only five trials, or 2% of the time). When a child missed a trial, his or her overall mean visual fixation times to the match and nonmatch across the remaining test trials were substituted in that cell. Thus, each child contributed eight data points to the analysis: mean visual fixation times to the match and to the nonmatch for each of four pairs of test trials.

Four factors relating to order of stimulus presentation were counterbalanced: (a) the number of matches on a screen side (two on the left and two on the right); (b) the order of the matches (either left-right-right-left or right-left-left-right) depending on whether Tape 1 was shown in the left or right deck; (c) the order of mention of the two actions during the salience trial (i.e., the action that was to be the match was labeled first for two of the verb pairs); and (d) the member of a verb pair that was labeled as the match (17 participants were asked to find the match for walking, picking flowers, running, and skipping, and 15 were asked to find the match for the mates of these verbs, or dancing, shoveling, rolling, and hopping).

Results

Test of Stimulus Salience From the Salience Trials

We predicted that the members of a stimulus pair would be of equivalent salience because the videotapes were constructed with this factor in mind. Stimulus salience was assessed by comparing visual fixation times during the salience trials when the linguistic stimulus was considered neutral because it did not exhort the child to look at either screen. Each child's mean fixation times on both salience trials for each member of a verb pair were entered into a three-way mixed analysis of variance with the factors of sex (2), verb pair (4), and match versus nonmatch (2). Although the linguistic stimulus during the salience trial did not direct children to watch either screen, the event that was to be the match during the test trials was taken to be the match for this analysis; the event that was to be the nonmatch in the test trials was taken to be the nonmatch.

No significant main effects or interactions resulted. Therefore, there were no a priori preferences for one verb or another in any pair. Table 3 presents these means.

Could Children Find the Matching Screen From Point-Light Displays?

To answer the main question addressed in the research, we conducted a three-way mixed analysis of variance with the same factors as the one described above. The only significant effect was the main effect of match, $F(1, 3) = 48.81$, $p < .01$. The mean visual fixation time to the match was 3.36 s, and that to the nonmatch was 2.29 s. Similar results were evidenced by the majority of the children. Out of 32 subjects, 29 (91%) had mean visual fixation times in favor of the match. Given the absence of an interaction with verb pair, this result indicates that children were able to find the match within each of the four verb pairs. Table 3 also presents these means.

Did the Pairs of Displays Differ in Similarity and Do These Ratings Predict the Mean Visual Fixation Scores to a Pair of Actions?

Two of the pairs were considered to be highly similar (picking flowers vs. shoveling and hopping vs. skipping), both receiving scores above 5 on the 7-point scale; the other two pairs (walking vs. dancing and running vs. rolling) received scores below 4, indicating that adults did not consider them to be that similar. Table 4 presents these means as well as the mean visual fixation difference scores (match minus nonmatch) for each pair of verbs. Though two of the verb pairs received higher similarity ratings, these ratings did not have any apparent effect on children's ability to find the match in these events.

Discussion

The purpose of this experiment was to determine whether young children could correctly extend familiar verbs in a comprehension task using actions depicted in point-light displays. When these motion-specified images were presented in the intermodal preferential looking paradigm (Golinkoff et al., 1987; Hirsh-Pasek & Golinkoff, 1996), 35–38-month-old children ($M = 37.0$ months) demonstrated extension of eight familiar verbs by watching the screen that matched the requested verb more than they watched the screen showing its mate. In the absence of a specific agent, and without contextual or featural information, children were apparently able to abstract the verbal essence of these displays and to

Table 3
Mean Visual Fixation Times (in Seconds) to the Stimulus Pairs During Salience and Test Trials

	Salience trials				Test trials[b]			
	Match		Nonmatch		Match		Nonmatch	
Verb pair[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Walking vs. dancing	2.74	1.46	2.82	1.55	3.67	2.29	2.22	1.20
Picking flowers vs. shoveling	2.49	1.15	2.84	1.29	3.31	1.69	2.37	1.50
Running vs. rolling	2.39	1.05	2.71	1.18	3.40	1.87	2.17	1.37
Skipping vs. hopping	2.73	1.49	2.53	1.46	3.16	1.56	2.05	1.71

[a] The verb requested in each pair (i.e., the match) was counterbalanced. [b] The match was watched more than the nonmatch for all test trials, $p < .05$.

Table 4
Mean Adult Ratings for Perceptual Similarity of Point-Light-Depicted Verb Pairs (Each Verb Compared With Its Mate)

Verb pair	*M*	*SD*
1. Walking vs. dancing	2.59	1.28
2. Picking flowers vs. shoveling	5.18	0.88
3. Running vs. rolling	3.18	0.95
4. Skipping vs. hopping	6.18	0.73

Note. Ratings were on a 7-point scale (1 = *not similar;* 7 = *very similar*). With the exception of Pair 1 versus Pair 3, multiple pairwise *t* tests revealed that all of the mean ratings differed from each other at $p < .01$.

extend familiar labels to these actions. Yet adult ratings of similarity within a verb pair revealed that discrimination between the actions in a pair was far from a simple task. Two of the verb pairs were rated as highly similar by adults. Thus, although adults perceived these verb pairs as sharing many perceptual characteristics, young 3-year-olds had little apparent difficulty extending familiar verbs to them.

These conclusions can be offered only tentatively. After all, the use of point-light displays to study motion verbs is completely novel. For this reason, we deemed it necessary to conduct an additional experiment, using a different kind of task, to provide converging evidence for the claim that children interpreted these point-light displays as action events. In the next experiment, we asked whether another set of children would spontaneously produce the appropriate motion verbs when shown these displays. Notice that this task is considerably more demanding than the comprehension task posed in Experiment 1. In the comprehension task, the verb label is given and can help children interpret what they are seeing. On the production task used in Experiment 2, by contrast, children have only the relatively ambiguous visual information at their disposal, and they must generate a label for this visual display.

Experiment 2

The purpose of this experiment was to see if we could generate collateral evidence for the assertion that young children saw these point-light displays as instances of particular actions. To this end, we showed the point-light displays to a different group of youngsters to see if they could provide names for these displays with minimal prompting. If young children were indeed capable of interpreting these displays as the motion events they were meant to depict, we would be in a better position to interpret the significance of the first experiment's results.

On the other hand, for us to be confident that children were seeing these displays as motion events, it was not necessary that the children generate names for the exact actions filmed. It is conceivable that a number of different verbs might be generated for these minimal displays. What would not be acceptable, however, would be a large number of "I don't know" responses or responses that did not mention motion. Furthermore, on the basis of the rationale that if adults could not generate names for the displays, children would not be able to either, we also tested a group of adults on the same task, asking them to generate verbal responses in the same way. From our own experiences in watching these movies, as well as from pretesting, we predicted that participants would be able to generate motion verbs as labels for these displays.

Method

Participants

Of the 38 children who participated, 29 (16 boys and 13 girls; age range = 38.13 to 45.10 months; M = 42.49 months) comprised the final sample. Data were excluded for 3 children who failed to produce a description for more than three of the actions and for 6 children who refused to talk. All children were native English speakers. The five adults asked to generate names for the actions were college students. Another set of 15 college students was asked to rate the children's descriptors of the point-light actions on a scale from 1 (*inappropriate)* to 7 (*highly appropriate*).

Stimulus Videotapes

For the purpose of label elicitation, the eight human point-light displays used in Experiment 1 (see Table 1) plus the point-light image of the cat walking were assembled so that the actions appeared one at a time, for 12 s each, followed by 5 s of blank screen. Two different random orders of the eight events were created.

Procedure for Verb Elicitation

The child sat on the parent's lap in front of a single, 32-in. (81-cm) television screen. The parent was asked not to say anything. Training began when the experimenter told the child that they were going to see a fun videotape and that the child would be asked to tell what he or she saw. After a single viewing of the cat display, the experimenter paused the VCR and prompted the child a maximum of three times for a label for the action. First, the child was asked a question that lent itself more readily to a noun than to a verb response: "Can you tell me what that was?" If the child produced only an object label, such as "a cat" or "a doggie," the experimenter asked, "What was the cat/doggie doing?" which usually led the child to produce the name of an action. Alternatively, if the child first said something such as "dots" or "snow," the experimenter labeled the cat for the child and encouraged the child to label the action. If the child did not produce an action label after being probed twice, the experimenter asked, "Was the cat walking?" In practice, most children responded successfully to the first, noun-biased question and seemed to readily grasp the idea that they were to produce a label for the moving dots.

For testing, the same basic procedure was repeated for each of the human actions except that the experimenter did not produce any description if the child did not. The first question asked of the child was "Can you tell me what that was?" If the response mentioned only an actor, the child was asked, "What was he/she doing?" Finally, some children acted out what they saw. They were then asked, "Tell me what you're doing." If the child failed to produce a description for the action after being probed twice, the experimenter went on to the next action. In practice, most children readily responded with an action term after the first question. The experimenter made only neutral comments in response.

Procedure for Collecting Adult Ratings of Children's Responses

We assumed that there might be several verbs that could name each of these displays (e.g., dancing might be called "dancing," "twisting," and "exercising") and that young children might not be quite as precise as adults in using the verb for the action that was actually filmed. Adults were asked to rate only the portion of children's responses that mentioned motion (93.6%) on a 7-point Likert scale. Adults were tested in groups and shown each point-light display in one of the same two orders used for the children. They were told that they would be seeing a series of videos and would be asked to rate the appropriateness of a set of possible responses used to describe them. They were given booklets, with one page for each action, that were assembled to correspond to the order of the actions they would see. Each page had the rating scale at the top (with 1 = *very inappropriate* and 7 = *very appropriate*) and a list of children's actual responses interspersed randomly with the same number of foils. The foils were selected from the responses children offered for other actions. Because each of the eight actions had a range of from 4 to 15 unique responses and was paired with the same number of foils, adults rated different numbers of responses for each action. For example, "running" engendered four different responses from children, so the page corresponding to "running" had the children's four actual responses ("running," "walking," "jogging," and "marching") interspersed with four foils ("rolling," "bending," "digging," and "sipping") to make eight responses for the adults to rate.

After eliminating the responses that did not mention motion, we found that children gave 103 unique responses to the eight actions. To create the response sheets for the adults, we reduced these to 78 responses, collapsing across verb objects (i.e., "picking up snowballs" and "picking up stuff" were considered the same). Because each real response had a foil, adults rated a total of 156 items for the eight actions.

A high rating indicated that the adult thought a descriptor for walking ("walking") was very appropriate, whereas a rating of 1 indicated that another descriptor, say "sipping," was inappropriate. If children interpreted these displays correctly, the foils for each action should have received significantly lower rankings of acceptability than the real responses children generated. Adults took about 10 min to complete this task.

Results

The five adults tested in the same manner as the children had no difficulty generating the actual verbs whose actions were depicted in these point-light displays. Each adult was 100% correct, naming each of the eight actions correctly. This result indicated that children might well be able to generate an acceptable range of verbs to describe these actions let alone realize that they were actions.

Children's responses were divided into two categories: relevant responses, which included any reference to motion, and irrelevant responses, which included words or phrases that did not indicate motion, such as "dots." The first significant result was that the majority of children's responses (93.6%) were relevant. This indicated that children were seeing motion events in these point-light displays. Furthermore, 22 (77%) of the children were able to produce verbs for all eight actions. Equally impressive is the fact that most children produced labels for all actions after being asked the first question, "Can you tell me what that was?" The next analysis allowed us to evaluate whether the motion events children reported bore any resemblance to the motion events portrayed. In other words, would adults find children's verbal responses to be appropriate to the displays?

Importantly, and as predicted, adults' ratings for the children's responses across all the actions (M = 3.80, SD = 2.10) were significantly higher than their ratings for the foils (M = 1.45, SD = 0.84), $F(1, 155) = 84$, $p < .001$. The foils were rated as mainly inappropriate, whereas children's actual responses across all actions were rated roughly at the midpoint of the scale. Thus, although children were previously unacquainted with point-light displays, their responses were judged by adults to be generally appropriate descriptors of the actions.

The next analysis considered children's responses by individual action. On the basis of the adults' ratings, children's responses for most of the actions were indeed considered appropriate (see Table 5). Ratings greater than or equal to 5 were considered highly appropriate; ratings of 3–4, appropriate; and ratings of 1–2, inappropriate. The mean ratings for only a single verb (shoveling) fell into the inappropriate range (M = 2.69). For the rest of the verbs, mean ratings ranged from 3.17 (dancing) to 4.77 (walking). Examples of responses raters considered highly appropriate were "walking" for the point-light display of walking and "running" for walking. Examples raters considered appropriate responses were "person bending up and down" and "exercising" for the point-light action of shoveling; examples coded as inappropriate responses were "bouncing legs" for the action of running and "running" for the action of shoveling.

Within children's responses that were considered appropriate, we next asked what proportion of children's responses were considered highly appropriate (i.e., ≥ 5 on the 7-point scale). Of the 93.6% of relevant responses that mentioned motion, fully 58% of children's responses across all verbs captured the action closely

Table 5
Number of Children Who Gave Highly Appropriate Responses and Mean Ratings for Responses

		Highly appropriate responses[c]		All responses	
Action	No. of children[a]/Total[b]	*M*	*SD*	*M*	*SD*
Rolling	10/25	6.33	0.70	3.87	2.00
Dancing	15/28	6.91	0.57	3.17	2.14
Picking flowers	10/25	6.41	0.68	4.31	2.22
Running	18/27	6.74	0.07	4.65	2.56
Walking	26/29	6.97	0.28	4.77	2.63
Hopping	16/27	6.78	0.70	4.10	2.42
Skipping	6/25	6.74	0.42	3.46	1.72
Shoveling	3/19	5.80	0.58	2.69	2.13

[a] Number of children who gave highly appropriate responses. [b] Total number of children who gave responses. Note that some children did not give responses for some actions. [c] Highly appropriate responses were those with ratings ≥5 on a 7-point scale.

(M = 6.73, SD = 0.58). Interestingly enough, there was a significant correlation between the number of children who produced highly appropriate responses and adults' appropriateness ratings of those responses (r = .36, p = .001), suggesting that the more appropriate a response was from the adult standpoint, the more children gave that response. However, as Table 5 shows, few highly appropriate answers were given for the actions of skipping (24%) and shoveling (20%). Nonetheless, children produced a mean of 96.2% relevant responses for shoveling (e.g., "dancing" or "exercising") and a mean of 78.9% relevant responses for skipping (e.g., "running" or "jumping"), indicating that at least they saw motion in these displays.

Discussion

Experiment 2 was conducted to provide additional validity for the findings from Experiment 1. Results from Experiment 1 suggested that young 3-year-olds were capable of finding which of two simultaneously presented point-light displays corresponded to a verb they heard. However, converging evidence was needed to make sure that children were in fact seeing these point-light displays as familiar actions. We chose to examine children's ability to abstract the verbal essence of these displays in an even more demanding way by examining children's spontaneous verbal productions when they were presented with the displays one at a time. To evaluate whether the responses children offered were close to the mark—that is, close to the events that the actors portrayed—we asked adults to view these events and to rate the responses children offered on a Likert scale.

The results indicated that children did indeed see these point-light displays as depicting motion. Almost all (93.6%) of their responses used motion descriptors. Furthermore, the majority of their motion descriptors were considered either appropriate or highly appropriate, according to the adults' ratings. This means that when children see these displays, they can indeed generate motion verbs (for at least six of the eight actions shown) that are either identical or close to the actions portrayed. The degree of accuracy that 3-year-olds realized is quite impressive for two reasons. First, point-light displays are nothing but lights moving around against a black background. Second, children at this age are not known for their ability to generate verbal responses in ambiguous situations. It is worth noting that production of verbal labels is a considerably harder task than the comprehension of these labels (Hirsh-Pasek & Golinkoff, 1996). In a comprehension task, the verb is given to children and can therefore serve to help them interpret what they are seeing. In a production task, children have only visual information to go on, and this fact seems to result in less precise interpretations of the scene. The fact that 3-year-olds perceived movement— and, in particular, the fact that they generated a constrained set of motion verbs that were highly similar to the actual actions portrayed—lends validity to Experiment 1. It appears that when children watched the matching screen in Experiment 1, they did so because they were actually able to locate a particular action as requested by the linguistic stimulus. In Experiment 1, children did not have to generate a motion verb; they only had to comprehend the verb offered and watch the action in a pair of displays that corresponded to that verb.

General Discussion

To use verbs flexibly, children must be able to map verb labels onto categories of action. Thus, the word "running" applies equally well to Grandma and to Carl Lewis. As this example illustrates, extension along action categories appears to be a more challenging task than the extension of object labels. In this article, we have suggested that verb extension proceeds as children abstract the invariant semantic components that describe that action (its verbal essence). In an effort to probe children's verb extensions, we conducted two experiments using actions portrayed in point-light displays. The outcome of the first experiment suggested that children were capable of recognizing motion events in point-light displays. When shown two simultaneously presented point-light actions, children were able to indicate which one matched the verb that was being requested. The second experiment provided converging evidence for the findings from Experiment 1. Experiment 2 made the task more difficult by asking children to generate their *own* labels for the actions transpiring on the monitor. Adults' ratings of the children's productions indicated that the majority of their verbal responses named an action that was the same as the one actually filmed or a related action.

Children's ability to interpret point-light displays is intriguing. Previous research focused on infants' discrimination of canonical and noncanonical biomechanical images and adults' recognition of the familiarity and gender of point-light walkers. No work, however, has assessed the appropriateness of these displays for young children let alone combined the stimuli with a language task. Yet the children in Experiment 1 successfully interpreted these point-light displays. Thus the results go beyond prior work and establish that the images have meaning for young children. In this first experiment, to ensure that the verbs depicted in the displays were truly familiar to the participants, we used 3-year-olds. If the children had failed to find the match, we would not have known if their failure was due to a lack of knowledge of the particular verbs or to the children's inability to perceive anything meaningful in point-light images. Given that this demonstration of the utility of point-light displays was successful, we are now in a position to extend this technique to the study of novel verb learning in young children. Indeed, research being conducted in our laboratories is already exploring this approach. Previous research on verb learning may have been hindered (Tomasello & Merriman, 1995) because the agents, locations, and instruments were so salient in other experimental methodologies that they may have overshadowed the action. By using the point-light technique, there is reason to believe that we can more fully understand the type of information that underlies children's verb concepts.

Our results suggest that by 37 months of age, children seem capable of extending their known verbs to highly unfamiliar depictions of the actions they represent. Golinkoff, Jacquet, Hirsh-Pasek, and Nandakumar (1996) gave us reason to believe that verb extension would be quite robust by around 3 years of age. In their study, drawings of Sesame Street characters performing both familiar (e.g., eating) and novel (e.g., performing an arabesque) actions were shown to 34-month-old children. Children had no difficulty selecting the familiar action (out of drawings of four actions) when asked questions such as "Where's eating?" Impressively, they also mapped the novel verbs to the novel actions even though they were not explicitly told which action the novel verb described (see also Merriman, Marazita, & Jarvis, 1995). Thus, when asked to "Find daxing!" instead of selecting one of the three familiar actions, they chose, at a rate above chance, the character performing the novel action. Children were also able to extend the novel verb to yet another character performing the same action on the next trial even in the presence of possibly confusing foils. That is, they selected a *different* character "daxing" when they were again asked, "Where's daxing?" even in the presence of yet another unnamed, novel action. They also did not select the character that had performed the action on the prior trial, as would have been predicted had they relied on the agent of the action as the basis of extension. In sum, we predicted that children would be able to extend familiar verbs in the point-light study because in the study by Golinkoff et al. (1996), the children were able to extend a novel action word to yet another exemplar after only a single exposure to a two-dimensional representation of it.

Because Golinkoff et al. (1996) used drawings, however, the basis for children's verb extensions was still unclear. In these static images, events are no longer dynamic sequences but are artificially "frozen" in time and space. Thus, when drawings are used to present actions, they may paradoxically have more in common with object rather than with verb representations. Point-light displays, which permit the use of dynamic events, allowed us to create a stringent test of the hypothesis that children can readily extend familiar motion verbs. Because point-light displays contain so little explicit information, children had little to rely on other than the inferences they could make from the verbal essence—that is, the manner, the path, and so forth. This research opens the way for a more in-depth study of the concept of verbal essence. At this point in time, there is no literature on when and how young children first abstract components of actions such as manner and path. These components undergird the formation of action categories that in turn become the foundation for verb meaning.

If children can extend familiar verbs in point-light images and even provide verbal labels for them, the next question is whether they can learn novel verbs using these displays. Maratsos and Deak (1995) would predict that they could not. In their view, unlike the way constraints or word-learning principles allow children to make a quick guess about the referent of a novel noun, children are not able to benefit from a reliable set of heuristics to gain a quick and accurate reading of verb meanings. Even "movie" verbs (e.g., jumping), they claimed, do not possess a sequence of perceptually definable motions, and "do not show the straightforward potential for perceptual parsing . . . that concrete objects with the basic whole object makeup afford the [child]" (p. 392). Preliminary data from Maguire, Hennon, Hirsh-Pasek, et al. (2001) suggest that, to the contrary, 18-month-olds seem capable of learning novel verbs in point-light displays and of transferring verbs learned in point-light displays to scenes of actual individuals performing the same actions. Once we establish that new verbs can be learned in point-light displays, we will be in a position to test which components of the verb (e.g., manner, path, overall shape, or some other factor) govern extension. Experiments focusing on extension would allow researchers to discover which semantic components have to be preserved for children to consider a new instance of an action as an exemplar of a particular verb (Behrend, 1990).

It is, of course, important to consider that the real world does not present events in point-lights. In addition, not all verb types can be illustrated through these displays. Mental verbs such as "see" or "think," for example, cannot be easily depicted through such action sequences. Nonetheless, there are a sufficient number of motion verbs, and they appear early enough in language development, to warrant their continued exploration through this technique. If children begin verb learning by bleaching the details out of action events (e.g., Mandler, 1992, 1998) or extracting the invariants from action scenes (Gibson, 1966, 1979), then the ability to abstract verbal essence might be available almost from the start. The verbal essence is the glue that holds the meaning of the verb together and allows the child to figure out what event he or she is seeing—apart from the particular agent, location, or instruments used. If this hypothesis is correct, it may be that younger children will have an easier time learning the names of novel point-light displays of action than will older children. Preliminary data from our laboratories suggest that this is so (Maguire, Hennon, Slutzky, Golinkoff, & Hirsh-Pasek, 2001). Older children, who have many competing verbs in their lexicons, may have more difficulty ascertaining which aspects of the novel motion event contribute to its verbal essence. For example, if one knows the verb "run," one may need more data before one can extract the fact that "skip" has a different manner from "run" even though both entail forward motion with the legs. This view leads to a paradoxical prediction: Older children may need more exemplars to map new motion events to verbs than may younger children in possession of fewer verbs in their lexicons. Perhaps young children at first learn novel verbs in a gestaltlike way, without concern for extracting the specific aspects of the action that give it its verbal essence. Perhaps only later, when yet another verb is heard to label a similar event, do they struggle to find the unique components. These hypotheses are currently being explored.

It is important to note that achieving insights into the verb-learning process is tantamount to achieving insights into the acquisition of grammar because so much of the sentence revolves around the verb (Bloom et al., 1975; Gleitman, 1990). Discovering how children extract, categorize, and label actions in their environment is crucial to unmasking how word-to-world mappings take place.

References

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, N. Akhtar, L. Bloom, G. Hollich, L. Smith, et al. (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). New York: Oxford University Press.

Behrend, D. A. (1990). The development of verb concepts: Children's use of verbs to label familiar and novel events. *Child Development, 61*, 681–696.

Bertenthal, B. I. (1993). Perception of biomechanical motions: Intrinsic image and knowledge-based constraints. In C. Granud (Ed.), *Carnegie Symposium on Cognition: Visual perception and cognition in infancy* (pp. 175–214). Hillsdale, NJ: Erlbaum.

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science, 5*, 221–225.

Bloom, L. (1993). *The transition from infancy to language: Acquiring the power of expression*. New York: Cambridge University Press.

Bloom, L., Lightbown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development, 40*(2, Serial No. 160).

Choi, S., & Bowerman, M. (1991). Learning to express motion events in

English and Korean: The influence of language specific lexicalization patterns. *Cognition, 41*, 83–122.
Frawley, W. (1992). *Linguistic semantics*. Hillsdale, NJ: Erlbaum.
Gentner, D. (1982). A study of early word meaning using artificial objects: What looks like a jiggy but acts like a zimbo? In J. Gardner (Ed.), *Readings in developmental psychology* (2nd ed., pp. 137–142). Boston: Little, Brown.
Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). New York: Cambridge University Press.
Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
Gleitman, L. (1990). Structural sources of verb meaning. *Language Acquisition, 1*, 3–55.
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*, 23–46.
Golinkoff, R. M., Hirsh-Pasek, K., Mervis, C. B., Frawley, W. B., & Parillo, M. (1995). Lexical principles can be extended to the acquisition of verbs. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 185–221). Hillsdale, NJ: Erlbaum.
Golinkoff, R. M., Jacquet, R., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles underlie verb learning. *Child Development, 67*, 3101–3110.
Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for lexical principles. *Journal of Child Language, 21*, 125–155.
Golinkoff, R. M., Shuff-Bailey, M., Olguin, R., & Ruan, W. (1995). Young children extend novel words at the basic level: Evidence for the principle of categorical scope. *Developmental Psychology, 31*, 494–507.
Hennon, E., Hirsh-Pasek, K., Golinkoff, R. M., Rocroi, C., Arnold, K., Maguire, M., et al. (2000). *From proper nouns to categories: Infants learn how words work*. Paper presented at the meeting of the International Conference on Infant Studies, Brighton, England.
Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.
Hollich, G. (1999). *Mechanisms of word learning: A computational account.* Unpublished doctoral dissertation, Temple University, Philadelphia.
Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review, 90*, 72–93.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201–211.
Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: MIT Press.
Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23(2)*, 141–221.
Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics, 21*, 575–580.
Maguire, M., Hennon, E., Hirsh-Pasek, K., Golinkoff, R. M., Slutzky, C., & Sootsman, J. (2001, November). *Mapping words to actions and events: How do 18-month-olds learn a verb?* Paper presented at the meeting of the Boston Child Language Conference, Boston, MA.
Maguire, M. J., Hennon, B., Slutzky, C., Golinkoff, R. M., & Hirsh-Pasek, K. (2001). *Infants' development of lexical categories: Moving from perceptual to social cues*. Paper presented at the meeting of the Society for Research in Child Development, Minneapolis, MN.
Mandler, J. M. (1992). The foundations of conceptual thought in infancy. *Cognitive Development, 7*, 273–285.
Mandler, J. M. (1998). Representation. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (5th ed., pp. 255–308). New York: Wiley.
Maratsos, M., & Deak, G. (1995). Hedgehogs, foxes, and the acquisition of verb meaning. In M. Tomasello, & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 377–404). Hillsdale, NJ: Erlbaum.
Marconi, D. (1997). *Lexical competence.* Cambridge, MA: MIT Press.
Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.
Merriman, W. E., Marazita, J., & Jarvis, L. (1995). Children's disposition to map new words onto new referents. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 147–183). Hillsdale, NJ: Erlbaum.
Nelson, K. (1996). *Language in cognitive development*. New York: Cambridge University Press.
Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
Sharon, T., & Wynn, K. (1998). Individuation of actions from continuous motion. *Psychological Science, 9*, 357–362.
Smith, C. A., & Sachs, J. (1990). Cognition and the verb lexicon in early lexical development. *Applied Psycholinguistics, 11*, 409–424.
Soken, N. H., & Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Development, 63*, 787–795.
Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 57–149). New York: Cambridge University Press.
Tardiff, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology, 32*, 492–504.
Tomasello, M., & Merriman, W. E. (1995). *Beyond names for things: Young children's acquisition of verbs*. Hillsdale, NJ: Erlbaum.
Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science, 7*, 164–169.

Received June 1, 1999
Revision received February 20, 2002
Accepted February 20, 2002 ■