Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Experimental Child Psychology 95 (2006) 27–55
Journal of Experimental Child Psychology
www.elsevier.com/locate/jecp

The use of social and salience cues in early word learning

Carmel Houston-Price [a,*], Kim Plunkett [b], Hester Duffy [c]

[a] *School of Psychology, University of Reading, Reading RG6 6AL, UK*
[b] *Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK*
[c] *Faculty of Education and Language Studies, Open University, Milton Keynes MK7 6AA, UK*

Received 30 August 2005; revised 20 March 2006
Available online 4 May 2006

Abstract

This article explores young infants' ability to learn new words in situations providing tightly controlled social and salience cues to their reference. Four experiments investigated whether, given two potential referents, 15-month-olds would attach novel labels to (a) an image toward which a digital recording of a face turned and gazed, (b) a moving image versus a stationary image, (c) a moving image toward which the face gazed, and (d) a gazed-on image versus a moving image. Infants successfully used the recorded gaze cue to form new word–referent associations and also showed learning in the salience condition. However, their behavior in the salience condition and in the experiments that followed suggests that, rather than basing their judgments of the words' reference on the mere presence or absence of the referent's motion, infants were strongly biased to attend to the *consistency* with which potential referents moved when a word was heard.

Keywords: Word learning; Reference; Salience; Gaze direction; Covariation; Preferential looking

Introduction

Infants and preschoolers show a remarkable ability to acquire new vocabulary. They have been seen to require only a few exposures to a new word to learn its meaning in

* Corresponding author. Fax: +44 118 9316715.
E-mail address: c.houston-price@reading.ac.uk (C. Houston-Price).

doi:10.1016/j.jecp.2006.03.006

experimental settings (e.g., Carey & Bartlett, 1978; Dollaghan, 1985; Woodward, Markman, & Fitzsimmons, 1994) and demonstrate striking rates of vocabulary growth in more naturalistic longitudinal observations (e.g., Benedict, 1979). These feats of development are even more impressive if one considers the potential hazards children face when learning new vocabulary. On hearing each new word, the appropriate meaning must be selected from an infinite number of logically possible alternatives (Quine, 1960). Couched in these terms, word learning should be an intractable problem, and vocabulary acquisition should be time-consuming and error prone. The fact that it is not is the Quinean conundrum.

One aspect of this conundrum has received considerable attention: how young children determine the *type* of word they are encountering. Such research has shown that children's initial assumptions about the meanings of words are highly constrained (e.g., Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989) and that the syntactic framework in which a word is provided also influences assumptions about its class (e.g., Gleitman, 1990; Waxman & Booth, 2001). However, although such cues might lead children to interpret a new word as a count noun referring to a whole object, for example, they cannot help children to disambiguate which *particular* whole object might be the referent. Constraints that bias children to apply new labels to items for which they currently have no name (e.g., mutual exclusivity [Markman, 1989]; N3C [Golinkoff et al., 1994]) come closer to solving this second aspect of Quine's problem, but such biases are of little use to very young word learners who have no labels for most of the objects in their environment.

As a result, researchers have investigated whether children are sensitive to cues to reference provided by the learning environment itself. It has been claimed that the majority of children's vocabulary is acquired in joint attentional interactions with adults (Bruner, 1983; Tomasello & Todd, 1983). One line of research, therefore, has investigated whether children use cues provided by their social partners, such as their gaze, gesture, emotion, and intonation, to solve the disambiguation problem.

The development of children's ability to use such sociopragmatic cues to determine the meanings of words is well documented. At 13 months of age, infants will learn a new label for an object only if clear behavioral cues link the speaker and the object (Woodward, 2000); if the speaker ignores the object but labels it when the infant attends to it, no learning occurs. By 16 months of age, infants learn the appropriate mappings if an adult deliberately waits until the infants look away from the target before looking toward it and labeling it (Baldwin, 1991, 1993). On hearing the new word, infants look at the speaker, notice the speaker's discrepant gaze direction, and attach the word to the object toward which the speaker was looking rather than the object to which they themselves were attending. Finally, by 18 months of age, infants spontaneously check the speaker for a range of intentional clues (both emotional and behavioral) and interpret new words in the light of this information (Baldwin, 1995; Dunham & Dunham, 1992; Moore, Angelopoulos, & Bennett, 1999).

The subtlety of children's ability to "read" social cues early during the second year presents the social–pragmatic approach as a strong candidate solution for the problem of identifying the specific referent of a new word. However, the earliest evidence of comprehension, at around 6 months of age (Tincoff & Jusczyk, 1999), precedes many of the developments in referential understanding that are important for the social–pragmatic solution. Bloom (2000) also pointed out that some autistic children who lack an understanding of referential intent can nevertheless learn words. As a result, some have argued for the existence of a more basic mechanism for identifying words' meanings.

According to the associative learning perspective, the process of mapping words to referents uses domain-general learning mechanisms. Ample evidence exists that word–object relations can be learned very quickly by young infants in the absence of any obvious indication of referential intent (e.g., Houston-Price, Plunkett, & Harris, 2005; Schafer & Plunkett, 1998). However, such studies typically present a single unambiguous referent for each new word and, therefore, tell us little about how a child might identify a word's reference in the more ambiguous circumstances in which words are more commonly learned. Associative learning, therefore, is accepted by many as the "glue" by which word–referent links are stored in memory after children have discovered the appropriate referent (using word-learning biases, syntactic or sociopragmatic cues), but it is not generally seen as a mechanism by which word–referent links may be discovered.

However, associative learning theorists argue that general learning mechanisms *are* sufficient to resolve the disambiguation problem, at least during the early stages of word learning (Smith, 2000; Smith, Jones, & Landau, 1996). Smith and colleagues claim that the problem might be solved by attending to the relations that naturally occur between words and their referents. Such relations might manifest themselves in a variety of ways. For example, the presence of a referent in child's environment might be correlated with the use of its label in everyday language; over multiple exposures to a word, the most likely referent could be computed. Two recent studies have demonstrated that infants as young as 21 months of age are indeed able to compute the referent of a word from the statistical structure of the environment in which the word is heard (Akhtar & Montague, 1999; Mather, Schafter, & Houston-Price, 2006).

A second possibility is that the referent of a new word might be indicated by its relative novelty or salience compared with nonreferents at the time the word is used. The logic of this approach is as follows. People are more likely to talk about objects or events that are salient or new to the conversational situation. Infants prefer to attend to salient and novel objects and events. Infants, therefore, are likely to be attending to the referent on hearing the new word. As a result, appropriate word–referent links are formed. Support for such a role for referent novelty was provided by Samuelson and Smith (1998), who found that infants mapped new words to contextually novel objects in preference to objects that had been seen in the same situation before (but for an alternative account of infants' behavior, see Diesendruck, Markson, Akhtar, & Reudor, 2004).

Surprisingly little research has explored whether information about a word's meaning can be obtained from the relative salience of potential referents despite the fact that the salience afforded by an object's movement has been shown to facilitate novel word learning (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). Those studies that have manipulated the salience of potential referents have revealed more about infants' weighting of salience and social–pragmatic cues when the two conflict than about the use of salience cues in isolation. For example, Hollich and colleagues (2000) provided infants of 12, 19, and 24 months of age with either coinciding or conflicting social and salience cues to the reference of a new label. Hollich et al used a new interactive preferential looking paradigm, in which real objects were labeled by a live experimenter during the training phase. In their Experiment 2, the experimenter looked at and labeled either the more interesting or the less interesting of two objects, as determined by infants' baseline looking preferences. Infants at all three ages were sensitive to the experimenter's gaze direction during training. However, when infants were asked to find the labeled object during a subsequent test phase, only the oldest group's behavior was affected by the gaze cue manipulation. That is, the general preference for the more salient object that was seen across all age groups and conditions was disrupted among the 24-month-olds who had witnessed the experimenter gazing toward and labeling the less salient object during training. This study suggests, therefore, that by the end of the second year, infants are aware of the relevance of both cue types in word learning and that learning can be disrupted when they conflict.

In a further series of studies in which the two potential referents were balanced in their inherent salience, Hollich and colleagues (2000) manipulated the experimenter's behavior during labeling to explore which cues were necessary for 12-month-olds to learn a new word. Results showed that gaze direction alone was insufficient to induce learning and that only when the experimenter also handled the object and provided either more tokens of the label or a longer labeling time did infants attach the label to the object. Twelve-month-olds, therefore, require considerable support to identify a new word's referent, perhaps needing both social (eye gaze) and salience (handling) cues as well as multiple tokens of the word and sufficient processing time. Previous research would suggest that slightly older infants should succeed at this task given only the gaze cue (e.g., Baldwin, 1991). However, less is known about the age at which object salience might provide a sufficient cue.

Only one study to date has set out to investigate the role played by object salience in disambiguating reference. In Moore and colleagues' (1999) study, infants were required to identify the referent of a new word produced by an experimenter who looked either straight toward the infant or toward one of two potential targets on either side of the experimenter. Salience was manipulated by rotating and illuminating one of the objects, allowing infants' use of the salience and gaze cues to be assessed in isolation as well as in coinciding and conflicting conditions. Both manipulations influenced infants' direction of attention during the training phase. At test, infants were required to select one object from an array of four objects (target, nontarget, and two additional distracters) as the referent of the trained word. Both 18- and 24-month-olds showed that they had learned the correct word–object mappings when the gaze cue was provided in isolation (and when salience and gaze cues coincided), confirming previous findings that infants are able to use social cues to assign reference by the middle of the second year (e.g., Baldwin, 1995). However, when gaze direction was neutral and only the salience cue was provided, only the older group mapped the new word to the moving illuminated object. Moore and colleagues attributed the older group's success in this condition to infants' understanding of the "referential force" of the adult's language. However, this referential force could not have revealed which object was the referent; only the perceptual salience of the target could have been responsible. This study, therefore, demonstrated that social cues to reference are not necessary at 24 months of age and that salience cues can also be used by infants of this age.

The 18-month-old group's failure to identify the referent in the salience condition led Moore and colleagues (1999) to claim that "for novel word learning to proceed at the younger age, there has to be referential behavior by the adult in the form of a gaze" (p. 67). However, it is worth noting that in this condition, rather than presenting the salience manipulation without any gaze cue at all, the experimenter looked straight toward the infant throughout. The younger group's failure to attach the label to the salient object, therefore, may have resulted from these infant's finding the experimenter's face more salient than the moving target. Alternatively, infants might have noted that the adult ignored the moving object while providing the label; therefore, they may have treated the situation as a conflict condition (as demonstrated by Woodward, 2000, and described earlier). It is possible, then, that the ability to infer words' meanings from the salience of potential referents may be present at 18 months of age or even earlier and that this sensitivity has not been detected at such ages for the methodological reasons outlined. A true test of this ability would eliminate the distraction presented by the experimenter and present the new label via a loudspeaker, as in the standard intermodal preferential looking paradigm. The experiments described here employ such a procedure to investigate whether young word learners are indeed sensitive to a salience cue to reference.

There is also a case for maintaining greater experimental control over the social cue provided in such investigations because the presence of a live experimenter during training introduces the possibility of experimenter bias. For example, in Hollich and colleagues' (2000) interactive paradigm, the experimenter cannot be blind to the condition under which the infants are being tested or to the side of presentation of the target on test trials. According to Hollich and colleagues, the standard preferential looking procedure prohibits the investigation of social cues. However, this is not the case. Infants' ability to follow head and gaze direction is not restricted to cues provided by a live experimenter; infants are also sensitive to the gaze direction of a realistic digitized image of a human face shown on-screen. For example, 3-month-olds (Hood, Willen, & Driver, 1998) and even neonates (Farroni, Massaccessi, Pividori, & Johnson, 2004) have been shown to rapidly shift their visual attention to peripheral stimuli in the gaze direction of a digitized image of a face. At 4 months of age, attention can be directed toward objects indicated by the gaze direction of an animated video of a face (Reid & Striano, 2005; Reid, Striano, Kaufman, & Johnson, 2004). Less is known about whether these abilities remain in older infants or indeed about whether such gaze cues are sufficient to promote word learning. The experiments reported here, therefore, explored whether infants would use a highly controlled gaze stimulus, rather than a live experimenter, as a cue to words' meanings.

A further methodological consideration relates to the number of novel words that infants are taught. When only one new word is provided during training, infants might select the target at test because it is the only object that was labeled during training (for a discussion of this issue, see Schafer & Plunkett, 1998). In Moore and colleagues' (1999) studies, for example, the experimenter labeled the target as "Dodo" nine times but provided no label for the nontarget, instead referring to it as "she" or "her" the same number of times. Moore and colleagues asserted that the experimenter's referential behavior toward the two objects was balanced and that infants' choice of the target at test could not result from differences in the extent to which it was the focus of the adult's referential behavior during training. However, this control fails to balance the adult's *linguistic* behavior toward the two objects; only the target was ever labeled with a new word. Therefore, infants' behavior at test, when infants were asked to "show me Dodo," might reflect their knowledge of which object had been labeled previously rather than their learning of the relation between the specific object and the label of interest, a subtle but important distinction. Designs that teach infants two new labels for two novel objects are the only means of circumventing this methodological problem. When both test objects have been labeled with new words previously, systematic selection of the target can be attributed with confidence to infants' learning of the desired mappings.

In this article, we describe four experiments that explored whether 15-month-olds, at the early stages of building a vocabulary and at a time when a sensitivity to social cues is developing, are responsive to tightly controlled social and salience cues to reference. Infants' ability to assign reference was investigated when each cue was presented in complete isolation as well as when the two cues coincided or conflicted. The experiments taught infants two novel word–image pairings using an adaptation of the standard intermodal preferential looking paradigm. In each experiment, two images were presented side by side on-screen and a salience cue, a gaze direction cue, or both indicated which image infants should map to a simultaneously presented label. As in Moore and colleagues' (1999) study, movement of the target image was used as the salience cue. This cue was presented in the absence of any interfering social cue; instead, recordings of the labels to be associated with each referent were played via a central loudspeaker. To ensure control over the social cue, a digitized video of a female face was displayed between the two referents; the face turned to look at the appropriate image when its label was presented, again via a loudspeaker.

In each experiment, infants saw a series of 12 experimental trials, each of which included a test phase followed by a training phase. In every test phase, the same two potential referents were presented as still images and one of the two novel labels was heard. In each training phase, the label was heard for a second time and the target image was disambiguated by its movement, by the gaze direction of the face, or both. It was hypothesized that infants would preferentially attend to moving or gazed-on images during the training phase of each trial and form mappings between these targets and the simultaneously presented labels. Infants' learning was expected to be revealed during the test phases of subsequent trials in their preferential looking toward named images. Finally, it was hypothesized that infants' learning would be robust when the social and salience cues coincided to indicate the same referent and that learning would be disrupted when the two cues conflicted in the referents they indicated.

Finally, it is important to note that the unusual design employed in our experiments, with the test phase of each trial *preceding* the training phase, allows confident interpretation of the results. During each trial, infants were tested initially on their understanding of one of the two new words; they saw the same two still images of the referents and heard one novel word. The word that infants heard was systematically counterbalanced so that each word was heard on six test trials; as a result, each image served as the target and distracter equally often—six times each. After the initial test period of the trial, infants' attention was directed toward the target image by one or more cues during the trial's training phase. If the two phases of the trial had been presented in a training–test sequence, infants' behavior during test phases might have been affected by their attention to the target during the preceding training phase. In our test–training design, however, this confound was removed. Similarly, it is important to note that trials were presented in random order. If this had not been the case, carryover effects from infants' attention during the training phase of one trial might have influenced their behavior during the test phase of the following trial. In our randomized design, however, there was no relation between the image that was the target on one trial and the image that was the target on the next trial; therefore, there was no possibility of such carryover effects systematically affecting infants' behavior. Thus, a reliable preference for the target or distracter image during the test phases of any experiment can be attributed to the learning of word–image associations on previous trials.

Experiment 1

Experiment 1 explored infants' ability to determine the referents of two novel labels when they were given a digitally recorded gaze cue to the appropriate mappings. Infants heard one of two novel labels and saw a female face turn and look toward one of two potential referents on her left or right side. In each recording of the gaze cue, the face looked briefly forward before turning to look to the left or right side because infants are less inclined to follow gaze during face-to-face interactions if the adult fails to first make eye contact (Hains & Muir, 1996).

Method

Design

Infants were presented with a series of 18 trials, of which 12 were experimental trials and 6 were "wake-up" trials. Experimental trials taught infants new word–image associations and assessed infants' learning of these associations; infants saw the same pair of images on every experimental trial. Wake-up trials maintained infants' interest by presenting pictures and labels that were expected to be familiar to them; infants saw one of two pairs of images on these trials. To ensure that wake-up trials were seen at regular intervals, trials were presented in three blocks, with each block consisting of two wake-up trials followed by four experimental trials. Trials were presented one after the other until infants tired of the experiment or had completed all three blocks.

All trials followed the same format, with the first half being a test phase and the second half being a training phase. During the test phase of each trial, infants saw two still images and heard a label. During the training phase, the label was heard a second time and a video of a woman's face appeared between the images, looked briefly toward the infants, and then turned to her right or left side. It was expected that infants would (a) preferentially attend to the gazed-on image during training phases, (b) learn associations between each novel label and the image toward which the adult gazed on hearing it, and (c) reveal this learning in preferential looking toward target images during the test phases of subsequent trials.

Participants

A total of 36 infants (18 boys and 18 girls) were recruited from a database of children whose parents had volunteered to participate in developmental studies. Infants had a mean age of 15 months 15 days (range = 14 months 23 days to 16 months 10 days). Infants were in good health, came from English-speaking families, and were accompanied by their primary caregivers (in nearly all cases their mothers).

The data from 9 infants were excluded from analyses because 5 infants refused to participate from the outset, 2 infants suffered from equipment failure, and 2 infants suffered from experimenter error. The remaining 27 participants were 13 boys and 14 girls.

Two further exclusionary criteria were adopted in this experiment and all subsequent experiments. If an infant looked away from the screen for longer than 1 s on more than half of the trials in any block, the infant's data from that block and any subsequent blocks were excluded from analyses. Second, the data from any infant who failed to provide at least two blocks of data (out of a possible three) were excluded.

All 27 infants who completed the current experiment satisfied these criteria and were included in analyses, with 26 infants providing data for all three test blocks and 1 infant completing two test blocks.

Materials

Visual stimuli

Infants saw six images of objects, all taken from the BBC children's television programs *Teletubbies* and *Come Outside*. Still images of each referent were created by capturing a video of the moving referent using Broadway Video Capture software and taking a snapshot from each video. The two images for which labels were to be taught were selected because they depicted objects for which no labels were expected to be known: a large windmill-like structure and a showerhead periscope (Fig. 1). These two images appear during every episode of *Teletubbies* and were familiar to the large majority of participating infants, according to parents' reports. The images have been used successfully as referents in previous word-learning research (Houston-Price et al., 2005). The two name-unknown images were shown as a pair during every experimental trial. Four further images of an airplane, a book, a dog, and a rabbit were chosen as wake-up stimuli because their names were likely to be known to infants, according to productive vocabulary norms provided by the LEX database (Dale & Fenson, 1993) and comprehension norms reported for the Oxford Communicative Development Inventory (CDI) (Hamilton, Plunkett, & Schafer, 2000). The book and airplane were always presented as a pair on wake-up trials, as were the dog and rabbit. All six still images were full color and measured 352×288 pixels ($\sim 28.5 \times 23.5$ cm on-screen). Images appeared in two fixed picture locations on the left and right sides of a large back-projection screen, with the two locations separated by 25.5 cm.

Videos of the head turn to the left and right sides were recorded against a white background using a digital video camera. A female actor faced the camera and then turned to the side as if something had attracted her attention. A number of head turns to each side were recorded, and the two that were best matched in speed of rotation were selected. Broadway Video Editing software was used to ensure that each video lasted 4000 ms, with the onset of the head turn at 40 ms and its completion within 1000 ms. Ulead Media Studio Pro 6.0 was used to resize the video clips of the face to 312×288 pixels so that they could be shown on-screen between the two still images. A snapshot from the video of the head turn to the left side can be seen in Fig. 2.

Fig. 1. Still images of the two name-unknown images used in experimental trials in Experiments 1 to 4.

Fig. 2. Still image taken from the video of the head turn to the left side, one of two videos used to indicate the target in Experiments 1, 3, and 4.

Auditory stimuli

The nonce words *shoofy* and *gopper* were used as novel labels for the following reasons. First, these words obey the rules of English phonology and are highly dissimilar to one another, sharing no phonemes in common. Second, bisyllabic words with a trochaic stress pattern are typical of spoken English, and the addition of a final vowel or schwa to monosyllabic English words is commonly seen in both infant-directed speech and infant productions.[1] Finally, these words have been successfully learned as labels for the same referents in previous research (Houston-Price et al., 2005).

Word samples were produced by a female, native British English speaker using infant-directed speech. Words were digitally recorded in a soundproofed room at 22,050 Hz using 16-bit mono sampling. Seven words were recorded several times each—*look*, *shoofy*, *gopper*, *aeroplane*, *book*, *dog*, and *rabbit*—and the best token of each was selected. Cool Edit 96 software was used to extract speech samples from the digital tape, and Goldwave was used to eliminate background noise and match the samples for amplitude. The labels were then digitally spliced to create six naming stimuli in the form "Look! … shoofy! … shoofy!" so that, in each naming stimulus, the label was heard twice, exactly 700 and 4700 ms after the onset of the word *Look*.

The assignment of novel words to novel images was counterbalanced between infants. Thus, half of the infants were taught that *gopper* referred to the windmill-like object and that *shoofy* referred to the periscope-like object, and half of the infants were taught the opposite pairings.

Procedure

The experiment took place in a child-friendly room with a booth designed for the intermodal preferential looking paradigm. Infants were seated on their mothers' (or other caregivers') laps approximately 1 m in front of a back-projection screen. The mothers lis-

[1] Note that in British English pronunciation (unlike American English), *gopper* contains a final vowel sound.

tened to music through headphones and were asked to keep their eyes closed throughout the experiment. Prior to the start of the first trial, infants' attention was drawn to the top center of the screen by a flashing red light controlled by the experimenter in an adjoining control room. Subsequent trials were initiated by the experimenter only when infants fixated the screen, and the light was used to redirect the infants' attention between trials when necessary.

Infants saw a series of up to 18 trials. The timing of every trial can be seen in Fig. 3A. At the start of each trial, infants were directed to look at the screen by the word *Look!* Infants then saw two still images on the left and right sides of the screen for the remaining 8 s of the trial. Halfway through this period, a face appeared between the images, looked toward the infants, and turned to look at the image on the left or right side. The target image, defined as the image toward which the head turned, was labeled twice during each trial: once 100 ms after the two images had appeared and again 100 ms after the face had appeared. The first 4000 ms in which the images were seen formed the "test phase" of each trial when infants' comprehension of the target word was assessed. The latter 4000 ms of each trial formed the "training phase" when infants were "taught" to associate labels with gazed-on images.

Trials were presented in three blocks of two wake-up trials followed by four experimental trials. During each block of wake-up trials, infants saw the airplane and book paired once and the dog and rabbit paired once. The target image during each wake-up trial was determined using a Latin square, and the side of presentation of the target was counterbalanced within each block. Order of presentation of the wake-up trials in each block was randomized by the presentation software.

During each of the four experimental trials of each block, the same pair of name-unknown images was presented (Fig. 1). The target image and its side of presentation were counterbalanced within each block so that each image was the target twice during each block: once when presented on the left and once when presented on the right. The order of presentation of experimental trials was randomized within each block. The experiment was terminated when all three blocks had been presented or when infants refused to continue with the experiment.

Scoring

Infants' looks toward the two images were recorded by cameras positioned above each of the image locations. The two recordings were spliced onto one video, and the lengths of infants' looks toward each image during the test and training phases of each experimental trial were scored twice offline using a button-press apparatus. The scorer was blind to the side of presentation of the target image. Intrascorer reliability was assessed by computing Pearson's correlation coefficients between the first and second sets of scored times for the left and right images for a random sample of 20% of infants. Mean reliability was $r = .94$ ($n = 6$, range = .93–.97). The first and second sets of scored times were then averaged for each participant.

Results

The proportion of time infants spent fixating the target image out of the total time spent fixating the target or distracter image was calculated for the test and training periods of each experimental trial—the 4 s before and after the onset of the head turn, respectively.

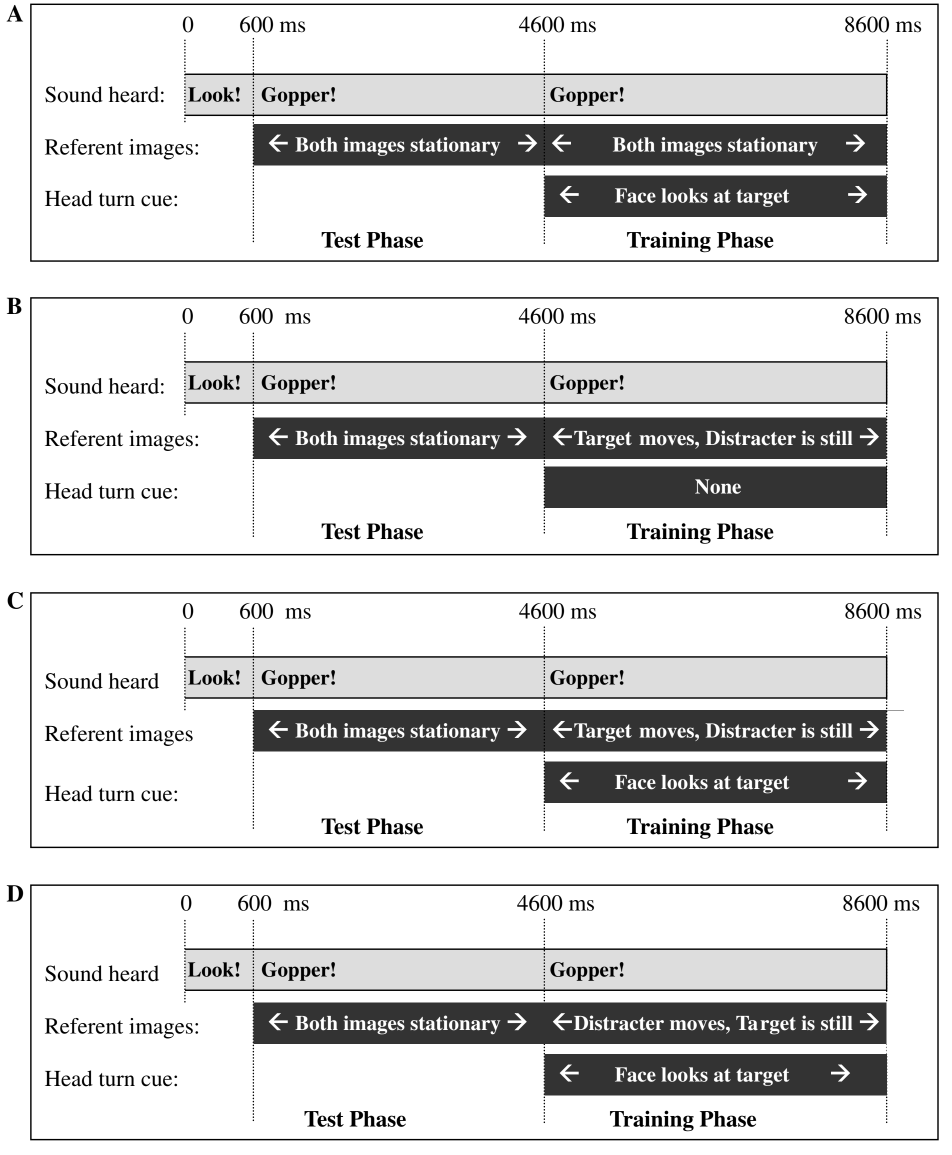

Fig. 3. Timing of each trial in Experiments 1 to 4. Auditory stimuli are shown in the lighter bars, and visual stimuli are shown in the darker bars. (A) Experiment 1, gaze cue only. (B) Experiment 2, salience cue only. (C) Experiment 3, gaze and salience cues coincide. (D) Experiment 4, gaze and salience cues conflict. Note that the image indicated by the gaze cue was designated the target in Experiment 4. *Gopper* is used as the target word merely as an example.

Analyses of infants' looking behavior during the test period of each trial exclude the first trial of the experiment because infants' behavior on this trial could not have resulted from the yet-to-be provided training. Analyses of looking behavior during the training phases of each trial include all 12 trials because the head-turn cue was available for infants to follow during every training phase. Block means were calculated by averaging the data across trials within each block, and overall means were calculated by averaging the data across all trials.

Arcsine transformed values were computed, as recommended for proportional data. Because virtually identical results were found using transformed and untransformed values, untransformed data are presented here and in all subsequent experiments for the sake of clarity.

Training phase

During the latter 4 s of each trial, the training phase, the face turned toward the target image while it was named. It was hypothesized that infants would follow the direction of the head turn and spend a longer time fixating the target image than the distracter image during this period.

As predicted, the majority of infants looked longer in the direction of the head turn ($n = 24$), with only 3 infants looking longer in the opposite direction. Of the time infants spent fixating either image, an average of 57.6% ($SD = 8.7$) was spent fixating the gazed-on image. This tendency was consistent across the three blocks (Block 1: mean = 58.7%, $SD = 13.4$; Block 2: mean = 62.3%, $SD = 16.7$; Block 3: mean = 55.9%, $SD = 14.3$), $F(2,50) = 1.15$, $p > .05$. A one-sample t test confirmed that, on average across all 12 trials, infants looked significantly longer at the gazed-on image than would be expected by chance (50%), $t(26) = 4.53$, $p < .001$. Infants, therefore, followed the gaze direction of the head turn shown on-screen.

Test phase

The next set of analyses explores whether infants' attention to the gaze cue during the training phase of each trial enabled them to learn associations between target images and simultaneously presented labels. It was hypothesized that if these associations were learned, infants would preferentially fixate the image matching the word heard during the test phases of experimental trials. Recall that during the test phase of each trial, no face was present and the target was indicated only by the word heard. The image that served as the target and the word heard on each trial were counterbalanced so that each word was heard equally often and each image served as the target and the distracter equally often, and trials were randomly ordered. Thus, systematic fixation of the target during the test phase could result *only* from the learning of word–image associations in previous trials.

The majority of participants ($n = 21$) indeed spent longer looking at the target image than at the distracter image, whereas only 6 infants preferentially fixated the distracter. Of the time infants spent looking at either the target image or the distracter image, an average of 52.2% ($SD = 4.4$) was spent fixating the target (Fig. 4). This tendency to fixate the target was consistent across the three blocks (Block 1: mean = 51.9%, $SD = 10.6$; Block 2: mean = 52.0%, $SD = 8.3$; Block 3: mean = 52.1%, $SD = 8.4$). A 2 (Word) × 3 (Block) repeated measures analysis of variance (ANOVA) found no effects of word, $F(1,24) = 0.00$, $p > .05$, or of block, $F(2,48) = 0.35$, $p > .05$; therefore, the overall mean was used as the measure of learning. Although infants' mean preference for the target appears to be a small deviation from chance, they in fact fixated the target significantly longer than would be expected by chance (50%), $t(26) = 2.56$, $p < .05$, with a reasonable effect size (Cohen's

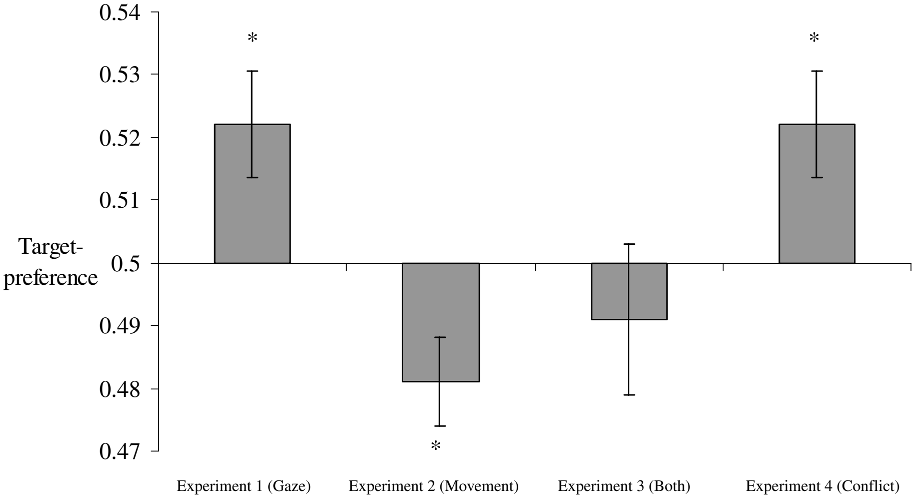

Fig. 4. Mean proportions of looking at the target during the test phases of experimental trials (excluding Trial 1) in Experiments 1 to 4, with standard error bars ($*p < .05$). Note that the image indicated by the gaze cue was designated the target in Experiment 4.

$d = .49$). In other words, infants used the information provided by the direction of the head turn to attach labels to gazed-on images during training, and this learning led them to preferentially fixate the named image during subsequent test phases.

Discussion

Experiment 1 shows, first, that infants as young as 15 months of age are sensitive to the gaze direction of a face presented on-screen. The development of the ability to follow gaze is well documented. It is known that by around 6 months of age, infants will follow the general direction of their mothers' head turns to the left or right side (Morales, Mundy, & Rojas, 1998) and can locate objects fixated by their mothers if they are salient and in the infants' visual fields (Butterworth & Grover, 1988, 1990). At around 12 months of age, infants begin to follow eye movements independently of any changes in head direction (Corkum & Moore, 1995; Lempers, 1979). Finally, between 12 and 18 months of age, infants' discrimination of the focus of adults' gaze grows in precision until they are able to locate objects outside their visual fields and detect the appropriate objects out of several possibilities (Butterworth & Jarrett, 1991). It is also known that very young infants are sensitive to the gaze direction of a digitized or animated face shown on-screen. Infants up to 4 months of age display shifts in visual attention in the direction of a recorded face's gaze (Farroni et al., 2004; Hood et al., 1998; Reid & Striano, 2005; Reid et al., 2004). The findings of the current experiment add to this literature by demonstrating that at 15 months of age, when the word-learning process is under way, infants reliably follow the head turn of a digitally recorded face shown on-screen.

Second, the data show that 15-month-olds are able to use such a prerecorded cue to determine the reference of new words. In the absence of a live interlocutor, infants attached new labels to images on the basis of only six trials pairing each new word with the head-turn cue. The literature concerning infants' use of social cues to learn new words reports significant developments in this domain between 12 and 18 months of age. At the beginning of the second year, infants require overt behavioral cues linking speaker and referent (Woodward, 2000). By the middle of the second year, infants spontaneously check the speaker's gaze and interpret new words accordingly (Baldwin, 1995; Dunham & Dunham, 1992). The current experiment further demonstrates that at 15 months of age, infants' knowledge of the significance of gaze direction is sufficiently robust that they will use a prerecorded cue in service of word learning. When the possibility that an experimenter might provide biasing or confounding nonverbal cues is removed, infants are able to attribute reference purely on the basis of the information given in the head turn.

Finally, previous research into infants' use of social cues to disambiguate reference has invariably concerned the ability to interpret cues provided by the *speaker* of a new label (e.g., Baldwin, 1993; Moore et al., 1999; Hollich et al., 2000). In the current experiment, the labels were presented via a central loudspeaker and the adult played the part of an interested "onlooker" in the labeling scenario. Although this should have made the task more difficult for infants, the fact that they were able to use the onlooker's gaze direction to form new word–object mappings corroborates the powerful nature of the gaze cue for young word learners.

Experiment 2

Experiment 2 investigated infants' use of image salience to determine which of two potential referents to associate with each of two novel labels. As in Moore and colleagues' (1999) study, salience was defined in terms of the movement of the target image because movement can be manipulated independently of other features of a stimulus and because movement is attractive to young infants and can facilitate word learning (Werker et al., 1998; but see also Houston-Price et al., 2005).

Method

Design

The design of this experiment was identical to that of Experiment 1 except that each word's referent was indicated by its movement rather than by an adult's head turn. The first half of each trial was the test phase when infants saw two still images and heard a label exactly as in Experiment 1. The second half of each trial provided the training; the target referent moved, the distracter image remained stationary, and the label was heard a second time. It was predicted that infants would (a) attend to the movement of the referent during the training phase, (b) learn associations between moving referents and their labels, and (c) reveal their learning by looking longer at named targets during the test phases of subsequent trials when both images were still.

Participants

A total of 32 infants (16 boys and 16 girls) were recruited in the same manner and tested in the same laboratory as in Experiment 1. No infant took part in more than one experiment reported in this article. Infants had a mean age of 15 months 8 days (range = 14 months 20 days to 15 months 28 days). After testing, the data from 2 infants were excluded from analyses: 1 due to experimenter error and 1 due to equipment failure. The remaining participants were 16 boys and 14 girls. Of these, 29 infants completed all three test blocks and 1 completed two blocks.

Materials

Visual stimuli

The two still images of name-unknown objects that were used in Experiment 1 were again used in this experiment (Fig. 1). "Moving images" of each object were created by capturing and editing a brief video of each moving object using Broadway Video Editing software. Each moving image consisted of a 4-s presentation of a still image of an object followed by a 4-s presentation of the same object moving. During the latter period, the windmill's sails rotated and emitted pink lights while the periscope rose from the ground and rotated toward the front. Stimuli were created in this way so that when a still image and a moving image were presented simultaneously, infants initially saw two stationary objects: one that remained stationary throughout the trial and one that began to move after 4 s and continued to move until the end of the trial.

Four new wake-up images were used in this experiment; these were captured from *Teletubbies* to maintain similarity with the images presented during experimental trials. Images of a baby, a bear, an elephant, and a frog were selected because their labels are learned early according to parental reports of receptive vocabulary on the Oxford CDI (Hamilton et al., 2000). Still and moving images of these items were created as described previously for experimental stimuli. The manner of movement shown by each wake-up image differed; the baby laughed, the bear skateboarded, the elephant walked, and the frog jumped. The baby and frog images always were presented as a pair during wake-up trials, as were the bear and elephant images. All images were full-color and measured 352×288 pixels ($\sim 28.5 \times 23.5$ cm).

Auditory stimuli

Auditory stimuli on experimental trials were identical to those used in Experiment 1. Naming stimuli were created for the four new wake-up words (*baby*, *bear*, *elephant*, and *frog*) in the same way as in Experiment 1.

Procedure

The procedure was identical to that of Experiment 1. Infants saw a series of up to 18 trials in three blocks of two wake-up trials followed by four experimental trials. The timing of each trial can be seen in Fig. 3B. Infants were directed to look at the screen 600 ms prior to the presentation of the visual stimuli by the word *Look!* They then saw two images on-screen: one of which remained stationary for the 8 s of the trial (the distracter image) and the other of which was still for the first 4 s and moved continuously during the latter 4 s (the target). The target was labeled twice during each trial: once 100 ms after the visual stimuli appeared and again 100 ms after the onset of its movement. The first 4 s during which images were presented provided the test phase of each trial when comprehension of the target word was tested. The latter 4 s provided the training phase when infants were taught to associate target words with moving images. Trial presentation was counterbalanced and randomized as in Experiment 1.

Scoring

Looking times were scored and averaged as in Experiment 1. Mean intrascorer reliability for a random sample of 20% of infants was $r = .97$ ($n = 6$, range = .96–.98).

Results

Looking times were measured during the test and training periods of each experimental trial—the 4 s before and after the onset of the target's movement, respectively. The data were averaged as in Experiment 1.

Training phase

During the latter 4 s of each trial, the training phase, the target image moved and the distracter image remained stationary. It was hypothesized that infants would spend longer fixating the target image during this period. Indeed, all 30 participants showed a preference for the moving image. Of the time infants spent fixating either image, an average of 77.4% ($SD = 5.5$) was spent fixating the moving image. This tendency was consistent across the three blocks (Block 1: mean = 80.1%, $SD = 8.4$; Block 2: mean = 79.3%, $SD = 10.6$; Block 3: mean = 76.9%, $SD = 9.0$), $F(2, 56) = 1.13$, $p > .05$. A one-sample t test confirmed that, on average across all 12 trials, infants looked longer at the moving images than would be expected by chance (50%), $t(29) = 27.30$, $p < .001$.

Test phase

The next set of analyses explored whether infants' attention to moving targets during the training phase of each trial enabled infants to learn associations between these images and the simultaneously presented labels. It was hypothesized that if these associations were learned, infants would reveal their learning during the test phases of the experiment by preferentially fixating the named target. Note that during the test phase of each trial, both images were still and only the word heard indicated which was the target.

Infants did not behave as expected. The majority of infants ($n = 21$) looked longer at distracter images during test phases, with only 9 infants preferring to fixate target images. Of the time infants spent looking at either the target image or the distracter image, an average of 48.1% ($SD = 3.9$) was spent fixating the target (Fig. 4). Infants' tendency to fixate the distracter image was consistent across the three test blocks (Block 1: mean = 44.2%, $SD = 7.9$; Block 2: mean = 49.1%, $SD = 9.1$; Block 3: mean = 49.0%, $SD = 10.5$). A 2 (Word) × 3 (Block) repeated measures ANOVA found no effects of word, $F(1, 28) = 0.21$, $p > .05$, or of block, $F(2, 56) = 1.61$, $p > .05$; therefore, the overall mean was used as the measure of learning. Again, although infants' mean preference for the distracter appears to be a small deviation from chance, they spent significantly less time looking at the target than would be expected by chance (50%), $t(29) = 2.63$, $p < .05$, with a reasonable effect size (Cohen's $d = .48$). The direction of infants' preference was unexpected, but the systematic nature of their behavior shows that infants were not reacting randomly to the novel words during test phases and, therefore, that their behavior was driven by their learning of word–image associations on previous trials.

Discussion

In this experiment, infants preferentially attended to moving images during the training phase of each trial. This is not surprising given that infants are known to prefer moving stimuli to identical stationary stimuli from a very young age (Slater, Morison, Town, & Rose, 1985). Hence, movement is a highly effective manipulation of salience. Behavior during the test phase of each trial was more unexpected. Infants demonstrated their learning of word–image associations by looking significantly longer at distracter images than would be expected by chance. Before considering possible causes of the surprising direction of infants' looking preferences, it is important to emphasize that infants' systematic behavior could have arisen only as a result of learning during the training phase of each trial. Learning was assessed during the first 4 s of each trial when both images were stationary and the target was indicated only by the label heard. Both the image that was the target on any trial and its side of presentation were counterbalanced, and the order in which trials were presented was random. Any attentional bias toward one side of the screen or one particular image, therefore, would result overall in chance levels of looking toward the target and distracter images. The fact that infants' behavior was not at chance demonstrates that it was driven by the label presented. Moreover, the fact that infants were taught two word–image associations means that their systematic behavior during the test phase can be unequivocally related to the learning of specific word–image relations during training (cf. Hollich et al., 2000; Moore et al., 1999). Therefore, this experiment demonstrates that by 15 months of age, a sensitivity to the relative movement of potential referents can be used to form new label–image associations.

Although several studies have explored the utility of object movement in facilitating word learning in nonambiguous learning situations (e.g., Houston-Price et al., 2005; Werker et al., 1998), the literature concerning the disambiguating role of object movement is limited to a single study. Moore and colleagues (1999) reported that the ability to use movement to infer reference was present at 24 months but not 18 months of age, and they suggested that the utility of the salience cue, therefore, is discovered during the latter part of the second year. However, as was discussed earlier, infants who participated in Moore and colleagues' study were faced by an experimenter who gazed at the infants while the target object was animated, and this competing gaze cue may have interfered with younger infants' attention to the movement cue. In the current experiment, no such competing cue was present and infants as young as 15 months of age used the salience cue to determine the referents of the new words. Moreover, learning was accomplished rapidly and with ease on the basis of only six presentations of each label. Previous research using the same images and labels in an explicit labeling task showed that 18-month-olds were able to learn associations between them when provided with three trials pairing each word and referent (Houston-Price et al., 2005). In the relatively difficult learning environment provided by the current experiment, in which the referent of each label needed to be inferred, six learning opportunities per novel label were sufficient for infants at an even younger age.

Nevertheless, the direction of infants' looking behavior remains to be explained. Two conflicting explanations are considered here and explored in the experiments that follow. The first is based on the assumption that infants learned the expected associations between labels and moving targets during the training phases of experimental trials. If this assumption is correct, it follows that during the test phase of subsequent trials infants revealed their learning by looking away from the image they believed to be the referent of the word heard and looking toward the distracter image. Such a preference for the mismatching image might appear to be counterintuitive given the fundamental assumption that underlies the preferential looking paradigm—that infants seek out matching intermodal stimuli. Indeed, it has been claimed that infants *never* show a significant preference for the nontarget in this paradigm (Hollich, Hirsh-Pasek, & Golinkoff, 1998).

However, other similar paradigms look for evidence of cross-modal learning in preferences for mismatching pairings. For example, in Werker and colleagues' (1998) "habituation switch" task, infants are presented with a series of label–image pairs in which the same label is always paired with the same image. After habituation to these pairings has occurred, infants are presented with a "switch" or mismatching pairing, where a label is paired with the image with which it was previously *not* presented. If infants' looking recovers on this switch trial, the infants are assumed to have detected the mismatch and, therefore, to have learned the pairings during the habituation phase. Longer looking times to mismatching word–referent pairings in this paradigm are assumed to result from infants' known preference for novelty once they have habituated to a familiar stimulus (or, in this case, to a familiar pairing of stimuli).

In the standard intermodal preferential looking procedure, infants' comprehension of a word is assessed typically during only one or two presentations of the word and corresponding image, a design that provides no opportunity for infants to habituate to the stimuli. In contrast, in word-learning experiments such as those reported here, infants may be presented with multiple instances of each word–image pair, allowing the possibility that infants might habituate to these pairings and, over time, seek out novel pairings as a result (for a more comprehensive discussion of this possibility, see Houston-Price & Nakai, 2004). Thus, one account of infants' behavior in Experiment 2 is that infants learned associations between words and moving images during training, habituated to the pairings they had learned, and, as a result, fixated images that created novel label–image pairings (distracter images) at test.

A second account of infants' behavior in this experiment makes a very different claim. This account postulates that from the infants' point of view, infants were not fixating distracter images at test; rather, they were fixating target images. We hypothesized that infants would attend to the movement of one image and form an association between this moving target and the simultaneously heard label. However, infants may have followed an alternative strategy. If they treated the image that remained stationary throughout each trial as the target and attached labels to still images as a result, their behavior at test would reflect a straightforward preference for the image that infants thought was being labeled.

Why might infants have attached labels to stationary images in this experiment? The most straightforward explanation is that infants are biased to attach new words to still images. The literature suggests, however, that this claim can be rejected. Werker and colleagues (1998) found that the movement of a referent facilitated the learning of its label for infants at 14 months of age, a similar age to participants in the current experiment. Our own work (Houston-Price et al., 2005) found no such advantage at 18 months of age, but importantly, we also found no benefit for stationary referents. Therefore, it is very unlikely that infants would prefer to assign reference to still images in the current experiment simply because they were stationary.

A more plausible explanation is suggested by the design of Experiment 2, which on reflection pitted the salience of the moving target against a second associative learning cue, the covariation of words and referents over multiple exposures. Infants have been shown to attend to the probabilistic cooccurrence of novel words and objects or images when forming new word–referent mappings (Akhtar & Montague, 1999; Mather et al., 2006). For example, Mather and colleagues (2006) gave infants multiple trials in which a novel label was presented in combination with a triad of novel images, one of which was the referent. Over a series of such trials, 21-month-olds learned appropriate image–label mappings purely on the basis of the covariation between the two. Akhtar and Montague (1999) similarly demonstrated that attention to the element (shape or texture) that remained constant across multiple instances of a new label allowed 2- and 4-year-olds to discover the meanings of adjectives.

In the current experiment, similar attention to the precise covariation of auditory and visual stimuli could have led infants to attach labels to still images rather than to moving images. If both the test and training phases of each trial were considered as providing useful information about the words' reference, one image (the target) sometimes was still and sometimes was moving when its label was heard, whereas the other image (the distracter) was consistently stationary when the same label was heard. It seems feasible, then, that infants might have used the information provided by this covariation to attach labels to still images. If so, looking behavior during test phases would be commensurate with "target looking" so far as infants were concerned.

It is important to stress that on either account, infants used the *relative movement* of the two images to determine the referent of each novel word, but according to one account infants attached labels to moving objects, and according to the other account they attached labels to stationary objects. The following experiments tease apart these two explanations while exploring how social and salience cues might interact in the disambiguation of reference.

Experiments 3 and 4 investigated whether infants benefit from coinciding cues to reference and whether their ability to use a single cue in isolation (as demonstrated in Experiments 1 and 2) is disrupted by the presence of a second conflicting cue. Experiment 3 provided infants with coinciding gaze direction and movement cues; both social and salience cues pointed to the same image as the referent of the word heard. Experiment 4 provided infants with conflicting gaze direction and movement cues; social and salience cues pointed to opposing images as the referent of the word heard. In addition to exploring the effects of providing coinciding and conflicting cues, this design allowed us to distinguish between the two accounts given above of infants' learning in Experiment 2. If infants learned labels for moving images (as in the first account), they should show robust learning in Experiment 3, when they are provided with an additional coinciding cue to the words' reference, and their learning should be disrupted in Experiment 4, when they are provided with a conflicting cue. If, however, infants learned labels for still images in Experiment 2 (as in the second account), the opposite pattern should be found in the experiments that follow; infants should show strong evidence of learning in Experiment 4, when the social cue supports their notion that the stationary image is the target, and learning should be disrupted in Experiment 3, when the social cue points to the moving image as the target, conflicting with the referent indicated by the infants' preferred learning strategy.

Experiment 3

Method

Design

The design of Experiment 3 was identical to that of Experiments 1 and 2 except that infants were presented with two coinciding cues to reference during every trial: an adult's head turn toward the target and the target's movement. As in Experiments 1 and 2, the first half of each trial was the test phase when infants saw two still images and heard a label, and the second half provided the training when the target was indicated by both its movement and the adult's head turn. It was predicted that infants would (a) attend to the referent that moved and was gazed on during the training phases of each trial, (b) learn associations between referents and simultaneously presented labels, and (c) reveal their learning by looking longer at named targets during the test phases of experimental trials when both images were stationary.

Participants

A total of 36 infants (23 girls and 13 boys) were recruited in the same manner and tested in the same laboratory as in Experiments 1 and 2. Infants' mean age was 15 months 7 days (range = 14 months 22 days to 15 months 27 days). Infants were in good health and came from English-speaking families. Following testing, the data from 6 infants were excluded from analyses because 2 infants failed to complete sufficient trials, the data from 2 infants were lost due to experimenter error, and the data from 2 infants were lost due to equipment failure. The remaining participants were 18 girls and 12 boys. All 30 participants completed all three test blocks.

Materials

Visual stimuli

The still and moving images of the two name-unknown objects that were used in Experiments 1 and 2 were again used in this experiment (Fig. 1). The wake-up stimuli that were used in Experiment 1 (airplane, book, dog, and rabbit) were used in this experiment because the wake-up stimuli used in Experiment 2 were insufficiently salient when stationary. Moving images of the wake-up stimuli used in Experiment 1 were created exactly as in Experiment 2. The video clips of the head turn that were used in Experiment 1 were used again in this experiment.

Auditory stimuli

The same auditory stimuli used in Experiment 1 were used in this experiment.

Procedure

The procedure was identical to that in Experiments 1 and 2. Infants were presented with a series of up to 18 trials grouped into three blocks of two wake-up trials followed by four experimental trials. The timing of each trial can be seen in Fig. 3C. Infants were directed to look at the screen 600 ms prior to the presentation of the visual stimuli by the word *Look*. They then saw two pictures side by side on-screen; one remained stationary for the remaining 8 s of the trial (the distracter), and the other was stationary for the first 4 s of the trial but moved continuously during the latter 4 s of the trial (the target). As the target image began to move, the face appeared between the two images and turned to look at it exactly as in Experiment 1. The target was labeled twice during the trial: once 100 ms after the visual stimuli appeared and again 100 ms after the appearance of the face and the onset of the target's movement. All other details were the same as in Experiments 1 and 2.

Scoring

Infants' looking times were scored and averaged as in Experiments 1 and 2. Mean intrascorer reliability for a random sample of 20% of infants was $r = .96$ ($n = 6$, range = .94–.98).

Results and discussion

Looking times were measured during the test and training periods of each experimental trial—the 4 s seconds before and after the onset of the head turn and movement, respectively. The data were averaged as in Experiments 1 and 2.

Training phase

During the latter 4 s of each trial, the training phase, the target image moved and was gazed on by the face while the distracter image remained stationary. It was hypothesized that infants would look longer at the target during this period. As expected, all 30 infants showed a preference for the moving gazed-on image during the training phase. Of the time infants spent fixating one of the two images, an average of 77.0% ($SD = 7.0$) was spent fixating the target. This tendency was consistent across the three blocks (Block 1: mean = 81.3%, $SD = 7.3$; Block 2: mean = 75.9%, $SD = 12.9$; Block 3: mean = 75.4%, $SD = 15.2$), epsilon corrected average $F(1.33, 38.62) = 2.24$, $p > .05$. A one-sample t test confirmed that, on average across all 12 trials, infants looked longer at target images than would be expected by chance (50%), $t(29) = 21.10$, $p < .001$.

Test phase

The next set of analyses explored whether infants' attention to targets during the training phase of each trial enabled them to learn associations between these images and the simultaneously presented labels. It was hypothesized that infants would reveal their learning during the test phases of trials by preferentially fixating the image that matched the word heard. Note that during the test phase of each trial, both images were stationary and no face was present; only the word heard indicated the target image.

In this experiment, looking behavior was close to chance during test phases. Of the 30 infants, 14 looked longer at the target during test phases and 16 looked longer at the distracter. Of the time infants spent looking at either the target image or the distracter image, an average of 49.1% ($SD = 6.5$) was spent fixating the target (Fig. 4), and infants' preference was not consistent across the three blocks (Block 1: mean = 50.8%, $SD = 13.2$; Block

2: mean = 48.4%, $SD = 12.1$; Block 3: mean = 48.6%, $SD = 9.1$). A 2 (Word) × 3 (Block) repeated measures ANOVA found no effects of word, $F(1,29) = 0.01$, $p > .05$, or of block, $F(2,58) = 0.62$, $p > .05$; therefore, the overall mean was used as the measure of learning. A one-sample t test confirmed that infants' mean target preference did not differ from chance (50%), $t(29) = 0.80$, $p > .05$, and the effect size was very small (Cohen's $d = .15$). Therefore, no evidence of learning was found in this experiment.

Although it was hypothesized that learning would be facilitated by the provision of two coinciding cues to words' reference, no evidence of learning was found in Experiment 3. When both a social cue and a salience cue indicated the same referent for each novel label, infants failed to form word–image associations despite their successful use of each cue in isolation in earlier experiments. It is possible that infants failed for the reasons outlined following Experiment 2. If infants were inclined to attach new labels to the images that consistently covaried with those labels (still images) rather than to the images that inconsistently covaried (moving images), then, in the current study, when the face always gazed on moving images, the two cues would have provided infants with conflicting information and their failure to learn would be unsurprising.

Strong support for such a claim can come only from Experiment 4. If infants' learning is restored by pitting the social and salience cues against each other so that the face gazes on the still image during each trial, this would provide strong evidence that infants are biased to link words and referents on the basis of a relationship of precise covariation.

Experiment 4

Method

Design

The design of this experiment was identical to that of Experiment 3 except that the two cues indicated different referents during the training phase; the face gazed on one image while the other image moved. As in Experiment 3, the first half of each trial was the test phase when infants saw two still images and heard a label, and the second half provided the training when the referent was indicated by the two conflicting cues. Based on the results of Experiments 1 to 3, it was predicted that infants would learn associations between the gazed-on still images and novel labels and reveal their learning by looking longer at target images during the test phases of subsequent trials.[2]

Participants

A total of 32 infants (20 girls and 12 boys) were recruited in the same manner and tested in the same laboratory as in Experiments 1 to 3. Infants had a mean age of 15 months 5 days (range = 14 months 16 days to 16 months 15 days). All infants were in good health and came from English-speaking families. No infants were excluded from analyses. Of the 32 infants, 31 completed all three test blocks and 1 completed two blocks.

[2] For ease of explanation, in this study the gazed-on image is referred to as the target and the moving image is referred to as the distracter, but note that infants might attach the label to either image in this design.

Materials

The same stimuli that were used in Experiment 3 were used in this experiment.

Procedure

As in previous experiments, infants were presented with up to 18 trials grouped into three blocks of two wake-up trials followed by four experimental trials. The timing of each experimental trial can be seen in Fig. 3D. Infants were directed to look at the screen 600 ms prior to the presentation of the visual stimuli by the word *Look*. They then saw two pictures side by side on the screen; one remained stationary for the remaining 8 s of the trial, and the other was stationary for the first 4 s of the trial and moved continuously during the latter 4 s of the trial. As one image began to move, the face appeared between the two images and turned to look at the still image. A novel label was heard twice during each trial: once 100 ms after the visual stimuli appeared and again 100 ms after the appearance of the face and the onset of movement.

All other aspects of the procedure were the same as in Experiments 1 to 3 except that in this experiment wake-up trials could not follow the same format as experimental trials. If infants heard a known word (e.g., *dog*) when shown conflicting cues about which image was the referent (e.g., a moving dog vs. a gazed-on rabbit), they might learn to attend to one cue and ignore the other and behavior on experimental trials might reflect this task-specific learning rather than infants' spontaneous reactions to the cues provided. On wake-up trials, therefore, infants were presented with either the gaze cue or the movement cue. Each cue was provided once during each block of wake-up trials in random order.

Scoring

Infants' looking times were scored and averaged as in previous experiments. Mean intrascorer reliability for a random sample of 20% of infants was $r = .94$ ($n = 6$, range = .82–.98).

Results and discussion

For analysis purposes, the gazed-on still image was considered to be the target in this experiment, and the moving image was considered to be the distracter. Looking times toward target and distracter images were measured during the test and training periods of each experimental trial—the 4 s before and after the onset of the head turn and movement, respectively. The data were analyzed as in Experiments 1 to 3.

Training phase

During the latter 4 s of each trial, the training phase, the target image was gazed on while the distracter moved. It was hypothesized that infants would attend to both images during this time. In fact, all 32 infants showed an overall preference for the moving image, confirming that the movement manipulation was highly salient for infants. Of the time infants spent fixating either image, they spent an average of 68.7% fixating the moving image ($SD = 7.6$). This tendency was consistent across the three blocks

(Block 1: mean = 70.9%, $SD = 12.6$; Block 2: mean = 69.0%, $SD = 14.2$; Block 3: mean = 64.1%, $SD = 13.6$), $F(2, 60) = 2.62$, $p > .05$. A one-sample t test confirmed that, on average across all 12 trials, infants looked significantly longer at the moving image than would be expected by chance (50%), $t(31) = 14.00$, $p < .001$.

Test phase

The next set of analyses explored whether infants' preferential attention to moving images during the training phase of each trial caused infants to learn associations between labels and moving images or whether infants attached labels to the stationary gazed-on images, as was predicted on the basis of Experiments 1 to 3. Recall that in this experiment, the target refers to the image that was still and gazed on during training, not the image that moved. As in previous experiments, during the test phase of each trial, both images were stationary and only the word heard indicated which was the target.

In this experiment, the majority of infants ($n = 20$) preferentially fixated the target during test phases and 12 infants preferred to fixate the distracter. Of the time infants spent looking at either the target image or the distracter image, an average of 52.2% ($SD = 4.8$) was spent fixating the target (Fig. 4). Infants' tendency to fixate the target was consistent across the three blocks (Block 1: mean = 51.9%, $SD = 10.6$; Block 2: mean = 53.6%, $SD = 10.4$; Block 3: mean = 53.7%, $SD = 9.3$). A 2 (Word) × 3 (Block) repeated measures ANOVA found no effects of word, $F(1, 30) = 0.00$, $p > .05$, or of block, $F(2, 60) = 0.84$, $p > .05$; therefore, the overall mean was used as the measure of learning. Again, although infants' mean preference for the target appears to be a small deviation from chance, it was significantly greater than would be expected by chance (50%), $t(31) = 2.60$, $p < .05$, with a reasonable effect size (Cohen's $d = .46$). Therefore, the data show that in this experiment, infants attached labels to gazed-on still images, as predicted.

Thus, as in Experiment 2, despite infants' greater attention to moving images during the training phase of each trial, infants did not attach simultaneously presented labels to these images. Instead, infants attached labels to images that were gazed on and stationary. These findings support those of Experiment 1—that 15-month-olds are willing to use a prerecorded social cue representing an interested onlooker to determine the reference of a new word. Taken on their own, the results might further lead one to conclude that despite infants' interest in the movement cue, infants ignored this salience cue and paid more attention to the social cue when identifying the words' referents. However, such a conclusion would not explain infants' pattern of failure and success in Experiments 3 and 4, respectively. Instead, the data provide strong support for the hypothesis presented following Experiment 2; despite the success of the movement manipulation in attracting infants' attention, when it comes to learning new word–image associations, infants are more concerned about the consistency with which a referent's motion covaries with the presentation of the word. To recap, whenever a given label was heard in Experiment 2, one image (the distracter) was consistently stationary during test and training phases and the other image's movement was inconsistent (the target image was stationary during test phases and moved during training phases). In that experiment, infants favored the still image when forming new word–referent associations. In Experiment 3, when the gaze cue pointed away from the consistently still image, infants showed no evidence of learning. In Experiment 4, when the gaze cue coincided with the consistency of the still image, learning was robust. Therefore, the data are highly indicative that infants are strongly biased to attend to consistency in the visual scene when forming new word–referent links.

Finally, we explored whether learning was facilitated when infants were provided with two coinciding cues compared with when they were given a single cue to the words' meanings. A univariate ANOVA compared the mean level of target preference shown by infants during the test phases of Experiments 1, 2, and 4.[3] There was no significant difference among the three experiments, $F(2,86) = 0.06$, $p > .05$. Independent samples t tests confirmed that there were no differences in target preference between Experiments 1 and 4, $t(57) = 0.02$, $p > .05$, or between Experiments 2 and 4, $t(60) = 0.31$, $p > .05$. Therefore, these analyses provide no evidence that learning was more robust when infants were provided with two coinciding cues to each word's reference. Thus, when two cues coincide to indicate the same referent, their utility is not simply additive in nature.

However, it should not be concluded from this finding that when two cues are provided simultaneously, infants are able to attend to only one of them. If this were the case, when infants were faced with conflicting cues in Experiment 3, they should have attended to one cue and ignored the other. Instead, infants showed no learning at all despite their high level of attention to the moving image. This suggests that infants at the early stages of word learning follow a conservative strategy of assigning meaning only when there is no discrepancy between the cues available.

General discussion

The primary aim of this series of experiments was to examine infants' ability to use a social cue (gaze direction) and a salience cue (object movement) to discover words' referents when these were presented in the absence of additional confounding cues. Previous research has failed in particular to isolate infants' ability to use perceptual salience to disambiguate reference from the effects of secondary interfering cues. For example, Hollich and colleagues (2000) investigated whether infants would attach a new label to a gazed-on object when they were presented with two potential referents, one of which was inherently more interesting than the other. This design is unable to reveal whether infants are able to use referent salience to determine words' meanings. Moore and colleagues (1999) controlled object salience by manipulating the movement and illumination of one of two equally interesting referents; however, when infants were provided with this salience cue, the speaker of the new word looked directly toward the infants, presenting an interfering and potentially conflicting cue to the word's reference.

Experiments 1 and 2 reported here presented infants with social and salience cues in isolation. These experiments revealed that infants used both social and salience cues in the process of building a vocabulary but that they used the salience cue in a surprising way. In Experiment 1, 15-month-olds attended to the images indicated by the adult's head turns and attached simultaneously heard labels to these images. Although previous research has demonstrated the use of gaze direction at a similar age (Baldwin, 1991, 1995), our data provide the first evidence that infants will infer reference on the basis of the gaze direction of an adult who is neither physically present nor interacting with the infants. The replacement

[3] For these analyses, the target was defined as the gazed-on image in Experiment 1, as the still image in Experiment 2, and as the gazed-on still image in Experiment 4.

of the live experimenter with a prerecorded gaze cue eradicates both the possibility of the infants being presented with additional directional cues over and above the intended cues during training and the opportunity for unintended or covert cueing of the target at test. Considering these controls, infants' ability to use the gaze cue is even more impressive.

It should also be noted that in these experiments, the gaze cues were provided by a face representing an observer to the communicative situation rather than by the speaker of the new labels. Recent research suggests that young children are not restricted to learning words that are addressed directly to them; they are also able to determine words' meanings by "overhearing" a labeling conversation between two adults (Akhtar, 2005; Akhtar, Jipson, & Callanan, 2001). Such studies demonstrate that word learning does not require the traditional formula of triadic joint attention among speaker, infant, and referent. Non-reciprocal relations between the child and the speaker are adequate if both parties are able to attend to the same referent. The data reported here add to this literature by showing that still more complex relations between social partners allow reference to be disambiguated. Infants will attach a new word to a referent on the basis of an attentional response to the word provided by a third party in the interaction. As Akhtar and colleagues (2001) pointed out, much "real-life" learning is likely to occur by observing others using and reacting to language; therefore, the social structures that enable vocabulary to be learned in this manner deserve further research.

Infants' learning in Experiment 2 provides the first evidence that the relative movement of two images can be used by infants as young as 15 months of age to determine the reference of new words. Moore and colleagues (1999) investigated the same ability in 18- and 24-month-olds and found that the older group spontaneously attached new labels to the more salient moving objects, whereas the younger group did not. However, as was discussed earlier, for Moore and colleagues' younger group learning may have been disrupted by the presence of the experimenter's neutral gaze cue. Experiment 2 provides evidence that 15-month-olds use relative movement as a referential cue. However, it is important to be clear about what we mean by relative movement. In our experiment, it was not the case that infants attached each new label to the referent that moved when the word was heard, as we had expected. Rather, infants attached each label to the referent that was consistently stationary when it was heard despite their strong interest in moving images during training phases.

This finding was particularly unexpected given previous reports that an object's movement is, in some circumstances, a prerequisite to the learning of its label (Werker et al., 1998). If movement is such a powerful facilitator of learning, why was its effect not seen in Experiment 2? It has been argued that infants' pattern of behavior is consistent with a word-learning strategy of attending to the consistency of the covariation between specific word–image pairs. If infants attached each new label to the part of the visual scene that was consistently seen when the label was heard, (i.e., the still image rather than the moving image), their behavior at test would be expected. That such a strategy was in fact driving infants' behavior was explored in Experiments 3 and 4 when infants were provided with a gaze cue that either coincided with or conflicted with the movement cue. Infants showed no learning when the gaze cue pointed to the inconsistently moving image and showed reliable evidence of learning when the gaze cue pointed to the consistently still image. These findings, therefore, provide strong support for the claim that infants treated still images as targets in this series of experiments, and we would suggest that stimulus covariation was the disambiguating cue that enabled infants' learning.

A further test of this hypothesis would involve running a modified version of Experiment 2 in which infants are presented with one moving image and one still image during each training phase, as in the current experiment, but see two moving images during each test phase rather than two still images. If it is the consistency of each image's movement when the label is heard that underlies infants' learning, in such an experiment labels should be attached to moving images rather than still images and looking preferences at test should be the reverse of those seen in Experiment 2. Such a finding would rule out the possibility that infants simply prefer to attach labels to stationary referents and would provide strong confirmation of our claim that stimulus covariation plays an important role in early word learning.

Previous research by Akhtar and Montague (1999) and Mather and colleagues (2006) showed that infants are sensitive to the concurrent presence or absence of potential referents when inferring the meaning of a new word. In both of these studies, the relevant referent (whether an object or a property of an object) was present on every occassion that its label was presented; indeed, infants' learning was based on this relation. The current experiments add to our understanding of this type of learning by demonstrating the strength of infants' desire for consistency between words and their referents at the very early stages of building a vocabulary. Given a choice between attaching a label to an image displaying a highly salient motion property that covaried with the label 50% of the time and attaching a label to an image with a considerably less salient nonmotion property that covaried with the label 100% of the time, infants gave more weight to the statistics of the covariation cue. Further research is needed to explore the limits of infants' ability to exploit such a source of information about words' meanings and whether this ability changes with age. For example, do young infants require a perfect relation between the occurrence of a word and a referent for an association to be formed between the two? Are older infants more flexible in their requirements, allowing more probabilistic statistical relations between words and their referents?

As our understanding of the nature of such associative learning mechanisms improves, it is becoming increasingly clear that the relationship between infants' attention to the target during labeling and their learning of the word–referent association is a complex one. Our studies suggest that it would be misleading to assume that the amount of time infants spend attending to potential referents determines *which* referent they will attach to a simultaneously presented label. In each of Experiments 2, 3, and 4, infants showed a strong preference for the moving image during the training phase of each trial, but in none of these experiments did infants attach the new label to this image. This finding has parallels with reports that infants are able to avoid erroneous mappings when an experimenter purposefully labels a different object to the object to which infants are attending (Baldwin, 1991, 1993). Infants do not appear to be biased to attach labels to whichever referents they are fixating at the time the labels are heard. Instead, they stand by a set of preferred strategies for determining reference. At 15 months of age, these include attaching labels to referents that are present in a consistent form whenever the words are heard and close observation of a social partner's direction of attention.

Acknowledgments

The research described in this article was supported by an Economic and Social Research Council project Grant (R000223732) to the first and second authors and by a Medical Research Council research studentship to the first author. We thank the parents and toddlers who participated in these experiments as well as Emily Mather and two anonymous reviewers for their comments on an earlier draft of the manuscript.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science, 8*, 199–209.

Akhtar, N., Jipson, J., & Callanan, M. (2001). Learning words through overhearing. *Child Development, 72*, 416–430.

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: the role of cross-situational learning. *First Language, 19*, 347–358.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875–890.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418.

Baldwin, D. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum.

Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of Child Language, 21*, 85–123.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bruner, J. (1983). *Child's talk: Learning to use language*. New York: Norton.

Butterworth, G. E., & Grover, L. (1988). The origins of referential communication in human infancy. In L. Weiskrantz (Ed.), *Thought without language* (pp. 5–24). Oxford, UK: Oxford University Press.

Butterworth, G. E., & Grover, L. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 605–624). Hillsdale, NJ: Lawrence Erlbaum.

Butterworth, G. E., & Jarrett, N. (1991). What minds have in common is space: spatial mechanisms serving joint attention in infancy. *British Journal of Developmental Psychology, 9*, 55–72.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15*, 17–29.

Corkum, V. L., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61–83). Hillsdale, NJ: Lawrence Erlbaum.

Dale, P. S., & Fenson, L. (1993). *LEX: A lexical development norms database [computer program]*. Seattle: University of Washington, Department of Psychology.

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: an alternative account of Samuelson and Smith. *Developmental Science, 7*, 33–41.

Dollaghan, C. (1985). Child meets word: "Fast mapping" in preschool children. *Journal of Speech and Hearing Research, 28*, 449–454.

Dunham, P., & Dunham, F. (1992). Lexical development during middle infancy: a mutually driven infant–caregiver process. *Developmental Psychology, 28*, 414–420.

Farroni, T., Massaccessi, S., Pividori, D., & Johnson, M. (2004). Gaze following in newborns. *Infancy, 5*, 39–60.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language, 21*, 125–155.

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development, 67*, 1940–1950.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory. *Journal of Child Language, 27*, 689–705.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (1998). Introducing the 3-D intermodal preferential looking paradigm: a new method to answer an age-old question. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 355–373). Norwood, NJ: Ablex.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., et al. (2000). Breaking the language barrier: an emergentist model for the origins of word learning. *Monographs of the Society for Research in Child Development, 65*, 3, Serial No. 262.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*, 131–134.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant & Child Development, 13*, 341–348.

Houston-Price, C., Plunkett, K., & Harris, P. (2005). "Word-learning wizardry" at 1;6. *Journal of Child Language, 32*, 175–189.

Lempers, J. D. (1979). Young children's production and comprehension of nonverbal deictic behaviors. *Journal of Genetic Psychology, 135*, 93–102.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Mather, E., Schafter, G., & Houston-Price, C. (2006). *Object-labeled covariation: a cue for the acquisition of nouns?* Unpublished manuscript.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*, 60–68.

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, 21*, 373–377.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: implications for cognitive neuroscience. *European Journal of Neuroscience, 21*, 1763–1766.

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport, 15*, 2553–2555.

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: an alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94–104.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by 15-month-olds under tightly controlled conditions. *Child Development, 69*, 309–320.

Slater, A. M., Morison, V., Town, C., & Rose, D. (1985). Movement perception and identity constancy in the new-born baby. *British Journal of Developmental Psychology, 3*, 211–220.

Smith, L. B. (2000). Avoiding associations when it's behaviorism you really hate. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). Oxford, UK: Oxford University Press.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition, 60*, 143–171.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*, 172–175.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4*, 197–212.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology, 43*, 217–242.

Werker, J., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-olds. *Developmental Psychology, 34*, 1289–1309.

Woodward, A. L. (2000). Constraining the problem space in early word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 81–114). Oxford, UK: Oxford University Press.

Woodward, A., Markman, E., & Fitzsimmons, M. (1994). Rapid word learning in 13- to 18-month-olds. *Developmental Psychology, 30*, 553–566.