Infants track word forms in early word–object associations

Tania S. Zamuner,¹ Laurel Fais² and Janet F. Werker²

¹. Department of Linguistics, University of Ottawa, Canada
². Department of Psychology, University of British Columbia, Canada

Abstract

A central component of language development is word learning. One characterization of this process is that language learners discover objects and then look for word forms to associate with these objects (Menamara, 1984; Smith, 2000). Another possibility is that word forms themselves are also important, such that once learned, hearing a familiar word form will lead young word learners to look for an object to associate with it (Jusczyk, 1997). This research investigates the relative weighing of word forms and objects in early word–object associations using the anticipatory eye-movement paradigm (AEM; McMurray & Aslin, 2004). Eighteen-month-old infants and adults were taught novel word–object associations and then tested on ambiguous stimuli that pitted word forms and objects against each other. Results revealed a change in weighing of these components across development. For 18-month-old infants, word forms weighed more in early word–object associative learning, while for adults, objects were more salient. Our results suggest that infants preferentially use word forms to guide the process of word–object association.

Introduction

Learning words for objects is a complex task that involves, at the very least, the cognitive processing of object concepts, word form meanings, and paralinguistic and social cues. Much research has been dedicated to understanding how children learn to connect object concepts and word forms. For example, work has examined constraints on the hypothesis space for the meanings of new words (Landau, Smith & Jones, 1988; Markman, 1992; Waxman & Markow, 1995); the impact of social cues such as eye gaze and joint attention (Baldwin, 1993, 1995); and the contribution of semantic knowledge and the understanding of the structure of words (Clark, 1973, 1983, 1993). Other work has examined the acquisition of categories and their internal structure (Markman, 1989; Mervis, 1987), and the impact of linguistic labels on how objects are categorized (Fulkerson & Waxman, 2007; Ferry, Hespos & Waxman, 2010) as well as on the understanding of different lexical classes (Hall, 2009). Moreover, word learning requires more than just learning words for objects; it also involves the complex mapping between word forms and concepts that do not have overt representation in the world. Jusczyk (1997) described the potential relationships between two specific components of word learning: ‘Sometimes infants may have a meaning in mind that they attempt to find the right sound pattern for. On other occasions, they may store a sound pattern first and then look to link it to the appropriate meaning’ (p. 133). This research investigates a precursor to learners’ sophisticated word learning and examines how 18-month-old infants might weigh these two components of word learning, i.e. acoustic word forms and the visual depiction of objects, in an associative word learning task.

How does the child go about the task of beginning to assemble a lexicon? Language learners acquire a rich knowledge of the sound patterns of their native language(s) within the first year of life (Gervain & Mehler, 2010; Werker & Curtin, 2005), as well as extensive experience with the people, objects, actions, and events that make up their world. This knowledge and experience are gained in a context in which linguistic and object-relevant input are richly entwined. Researchers have suggested that linguistic labels enhance object category perception (Waxman & Markow, 1995; Waxman & Lidz, 2006, for a review; but see Sloutsky & Robinson, 2008),
and evidence is emerging that the opposite is true as well, namely, that the consistent pairing of an object with a sound supports the perception of phonetic distinctions (Yeung & Werker, 2009; Yeung, Chen & Werker, in press). Word learning requires the integration of infants’ understanding of both word forms and objects, and that integration likely develops through the incremental coordination of advances in both perception of the sound patterns of the native language and comprehension of the nature, behavior, and function of objects. What has not yet been addressed, and what we explore in this work, is whether infants rely to a greater extent on their representation of word forms or objects in the process of word–object association.

There appears to be no doubt that infants do attend to the word forms that exist in their environment. Studies from the field of infant speech perception have revealed that infants have a remarkable ability to recognize certain word forms at an early age, and in the absence of meaning or overt referents. By 4.5 months of age, infants prefer to listen to their own names over foils matched for stress pattern, demonstrating that they are capable of storing linguistic sound patterns (Mandel, Jusczyk & Pisoni, 1995). At 6 to 7 months infants listen longer to passages containing familiarized words like bike and feet, and do not mistake these words for very similar non-words such as gike (Jusczyk & Aslin, 1995). By 11 months, infants have encoded word forms in sufficient detail that they show a preference for listening to a proper over an improper pronunciation of a word (Swingley, 2005). Infants’ and young children’s sensitivity to a wide variety of phonological, lexical and semantic characteristics of word forms is evident in the effects these features have on word learning (Jusczyk, 1997; Stoel-Gammon, 2011; Storkel, 2009; Zamuner, 2009); for example, familiarity with the word forms and objects facilitates 14-month-olds’ ability to link phonologically similar word forms (ball and doll) to objects (Fennell & Werker, 2003). In addition, simply having the opportunity to segment word forms from a speech stream supports 17-month-old infants’ ability to associate these word forms with objects (Graf Estes, Evans, Alibali & Saffran, 2007). Infants clearly pay attention to and retain information about the word forms they hear. This growing knowledge of word forms may help in early word learning (Swingley, 2009).

The ‘flip’ side of word–object association is, of course, the object. Other research has focused on investigating the effects of infants’ understanding of the object on their ability to make word–object associations. Pre-exposure to the object allows 14-month-old infants to establish a detailed enough pairing that they can use a minimal change in the novel words (din and gin) in a word–object association task (Fennell, 2011). Similar results have been found for older children; pre-exposure to the objects increases the likelihood that older children will link a novel object and a novel label (Graham, Turner & Henderson, 2005). In this work, infants are asked to choose an object named by a novel word form from a pair of familiar and novel objects. Infants who have been exposed to the novel object prior to testing are more likely to choose that object.

The majority of studies discussed above that examine infants’ early word–object associative learning have used a ‘Switch’ task paradigm (a variant of the Visual Fixation Procedure) in which infants are habituated to two novel word–object pairings and then tested on their responses to a habituated word–object pairing and to a novel pairing of a habituated word and a habituated object (Werker, Cohen, Lloyd, Casasola & Stager, 1998; Werker & Fennell, 2008; Werker, Fennell, Corcoran & Stager, 2002). An increase in looking time, or dishabituation, to the novel pairing is interpreted as indicating successful word–object association. Depending on the study, novel pairings consist of a change in the word form, a change in the object, or a change in both the word form and object. Using the Switch task, Werker et al. (1998) found no significant differences in 8-, 10- and 12-month-old infants’ looking to these different types of switch trials. In Werker et al. (2002), analyses did not examine the type of switch (change in word form or change in object); therefore, one cannot determine whether infants’ dishabituation was driven primarily by their knowledge of the word form or of the object. Similarly, the object choice task used by Graham et al. (2005) does not reveal whether participants in the task rely more on the word form or object to succeed.

In order to investigate how infants process word forms and object components in word–object associations, the current study used an adapted version of the Anticipatory-Eye Movement Paradigm (AEM; McMurray & Aslin, 2004). The AEM paradigm is a relatively new procedure in the field of language development research, used to date in only a handful of published research studies, but it is particularly well suited to our research question. The AEM is a type of two-alternative forced-choice task: typically, the participant watches two stimuli, presented serially, that disappear behind an occluder, and can emerge from two possible locations. After repeated exposure to the stimuli, participants begin to make anticipatory eye movements to the location where the visual stimulus will emerge, indicating that they have learned the relevant associations. In their original study, McMurray and Aslin (2004) found that infants were able to generalize new tokens of trained words as indicated by their eye movements to the correct
side for words that changed in pitch, but not duration. Because the AEM paradigm allows the dissociation of two aspects of the stimuli at test, and provides the participant with the choice of tracking one or the other component, it enables the explicit investigation of the influences of specific, multiple aspects of the stimuli on infant processing.

Our research uses the AEM paradigm to probe the process of forming word–object associations. In our adaptation of the AEM, infants were taught two word–object pairings during an initial Pre-training and Training phase. The objects (Figure 1a and 1b) moved behind an occluder as the relevant novel word form was played, and emerged consistently from one or the other side of the occluder (Figure 1c). Participants’ anticipatory looking to the ‘correct’ side of the occluder (i.e. before the object emerged) during the Training phase was analyzed to determine whether they had learned that the two pairs of stimuli were associated with different sides of the visual display. During the Test phase, participants were presented with stimuli in which the pairings of the trained word form and object stimuli were switched. As in the Training phase, the object stimuli moved behind the occluder; however, during Test trials, they did not emerge. Because the trained word form–object pairings were decoupled at test, the Test stimuli consisted of an object that had been associated during training with emergence from one side of the display and a word form that had been associated with emergence from the other side. Participants’ looking to each of the two sides was measured. The decoupling of word form and object crucially allowed us to examine which of these two elements of the word–object pairing was the more salient to the infant. A tendency to look toward one side rather than the other indicated that the aspect of the stimulus associated with that side – word form or object – was more heavily weighed in this task.

Experiment

The aim of the current study was to investigate the relative contributions of word forms and objects to early word–object associations. Two groups of participants were tested: 18-month-old monolingual English-learning infants and a group of adult controls. Infants at 14 months succeed in pairing different sounding word forms with novel objects (e.g. *lif* and *neem*; Stager & Werker, 1997), and by 17–20 months, infants are proficient enough word learners to be able to use more difficult minimal phonemic contrasts (e.g. *bih* and *dih*) to pair novel word forms and novel objects in the Switch task (Werker et al., 1998). In addition, younger infants have been shown to perform successfully in the AEM task (Albareda-Castellot, Pons & Sebastián-Gallés, 2011; Hochmann, Benavides-Varela, Nespor & Mehler, 2011; Kovács & Mehler, 2009; McMurray & Aslin, 2004). We reasoned, then, that the performance of 18-month-old infants in our paradigm would reflect meaningful, robust word learning behavior and not the effects of task demands or age limitations. Eighteen-month-old infants are proficient word learners, likely to be able to demonstrate this ability in an AEM task, and they are just on the brink of an explosion in vocabulary acquisition. Thus, they are an important age group for understanding the proclivities that infants bring to word learning as they grow more accomplished in this skill.

One view of word learning is that infants apprehend a referent or object category and then seek a label (Mcnamara, 1984; for discussion, see Jusczyk, 1997). This would predict that infants in our study will track the object, and not the word form, in our AEM task. The opposite account – that infants hold onto a word form and seek a meaning for it (Jusczyk & Hohne, 1997) – would predict that the infants in our study will track the word form and not the object. In other work examining infants’ abilities to process the auditory linguistic and visual object components in a cross-modal perceptual, though not a word learning task, Sloutsky and Robinson (2008) found auditory linguistic dominance for infants at 10 months, but no dominance for either modality at 16 months. Based on this finding, we might predict that if only perceptual preferences guide performance, and not word learning biases, 18-month-old infants would not show a reliance on either the word form or object stimulus in our word–object association task.

![Figure 1](https://example.com/figure1.png)

Figure 1  *Objects used for word–object pairings: (a) crown, (b) molecule. Occluder at center of visual display (c).*

© 2014 John Wiley & Sons Ltd
Method

Participants

Two groups of participants were tested (n = 16 in each group): 18-month-old infants (8 females; mean age 18 months, 20 days; range 17.7 to 20.0 months) and university age adults (12 females). Infants were born into homes where they received at least 80% exposure to English. Fifteen additional infants were tested and not included in the analysis for the following reasons: not completing the experiment (7), equipment failure (1), experimental data error (1), parental interference (2), insufficient data on Test trials (3), and out of range of viewer (1). Parents were recruited through visits to local hospitals, and were compensated with a small gift (T-shirt, bib, etc.) for their participation. Adult participants were undergraduate students recruited through the Human Subjects Pool of the Department of Psychology, who received course credit for participation. All adult participants and the parents of all infant participants gave informed consent, and all of the requirements of the University of British Columbia Behavioral Research Ethics Board were strictly observed.

Stimuli

The auditory stimuli consisted of two phonetically different novel word forms, lif and neem, recorded by a female in infant-directed speech. The average lengths of the stimuli for lif and neem were 999 ms and 1052 ms, respectively. The object stimuli were two, colored, novel objects molded out of plasticine and photographed (‘molecule’ and ‘crown’, Figure 1a and b; Stager & Werker, 1997; Werker et al., 1998; Werker et al., 2002). An occluder was located in the middle of the visual display (Figure 1c). Each participant was trained on two word–object pairs (lif–molecule, neem–crown or lif–crown, neem–molecule) and at test, the pairings were switched (lif–crown, neem–molecule or lif–molecule, neem–crown, respectively). Word–object pairings were counterbalanced across participants, as was the side of the occluder with which each pair was associated (left or right).

Apparatus

Infants sat on their caregivers’ laps about 24 inches from a 17-in. color Tobii 1750 eyetracker in a softly lit, sound-attenuated room. The position of the monitor could be adjusted for each participant to ensure efficient capture of eye movements. Parents either closed their eyes or wore a sleep mask; this both prevented them from inadvertently responding to the displayed stimuli, and ensured that the eyetracker recorded the infant’s eye movements and not the parent’s. The experimenter observed the behavior of the participants from behind a black divider by means of a low-light video camera placed just below the Tobii monitor. The camera was also used to record all sessions.

Before the start of each infant’s session, the calibration routine provided by Tobii was run. Infants tracked the sequential presentation of an interesting attention-getter at five points on the monitor. The routine was run, if possible, until infants’ eye gaze fell within 0.5 degrees of all five central points of the attention-getter; this ensured accurate data collection. An HP pavilion computer running Clearview software (version 2.5.1, Tobii Technology) was used to present the experiment stimuli. Eye movement data were recorded as x-y coordinates of the infant’s point of gaze, at 50 Hz, and saved to a file for each subject. Stimulus start and stop times were also recorded.

Procedure

An adapted version of the AEM paradigm was used in the study (McMurray & Aslin, 2004). The experiment consisted of a Pre-training phase, a Training phase and a Test phase. On Pre-training and Training trials, the object moved slowly and continuously from side to side above the occluder, as five different tokens of the associated word form were played separated by pauses of approximately 750 ms. The onset of the first audio token occurred concurrent with the beginning of the visual presentation of the object; each subsequent audio token began during periods of smooth motion in the movement trajectory, i.e. the presentation of the word form was not aligned to the side-to-side change in direction of the movement of the object. After five presentations of the word form (after ~8200 ms), when the object was centered above the occluder, it moved downwards. There were four Pre-training trials, which were identical to the Training trials, except that the object appeared in front of the occluder. This was to illustrate the object’s path of movement. On the remaining 16 Training trials, the object moved behind the occluder, and was hidden for 1400 ms. Finally, the object reappeared from behind the occluder (on either the right or left side) and one more repetition of the word form was presented. The reappearances were held constant throughout the Training phase for each word–object pair, such that participants saw objects reappear from both sides, but a given word–object pair only reappeared from a single side.

Infants were familiarized with the word–object pairs in four Pre-training trials and 16 Training trials (10 trials for each pair). Trials were organized into blocks of four, with two trials of each word–object pairing in...
semi-randomized order within each block. After participants completed the Training phase, the Test phase began. There were eight Test trials, organized into two blocks of four trials. On Test trials, the word–object pairings were switched; infants saw two instances of each switched pair in each of the two test blocks. Test trials also began with the object moving from side to side above the occluder, while the (switched) word form was repeated. Then, the object moved downwards behind the occluder; however, unlike in the Training trials, in the Test trials the object did not reappear from behind the occluder, and the word form was not played again. Figure 2 gives a schematic of the experiment procedure. The current design differed from that used by McMurray and Aslin (2004) in that it included distinct Training and Test phases, and did not include reinforcing Training trials during the Test phase. Also, in our implementation, the object moved down behind the occluder, whereas in McMurray and Aslin (2004) it moved up.

Before all trials, a silent attention-getter in the form of a looming and withdrawing colored oval appeared at mid-screen to orient participants’ attention. New trials began when participants looked towards the attention getter. After every block of four trials, participants were presented with short video clips approximately 12 s in length in order to help maintain participant attention. The clips consisted of cartoons showing animated penguin(s) (‘Pingu’) participating in actions that moved them downwards on the screen (skiing down a hill, bobsledding), or a toddler participating in an action that moved him downwards on the screen (going down a slide). We reasoned that watching these actions might reinforce participants’ understanding of the novel objects moving downwards, even when masked by the occluder. The entire experiment lasted approximately 8 minutes.

**Analyses**

The duration of looking for each participant’s eye movements to the left and right side of the visual display were analyzed separately for Training and Test trials. Measurements for the trials were made during three time periods and over three areas of analysis.

**Timing**

Time Window 1 began at trial onset and continued until 200 ms after the object began to move downwards behind the occluder, that is, until 8400 ms into the trial. Time Window 2 began at the end of Time Window 1 and continued until 200 ms after the object first started to appear from behind the occluder on the Training trials or would have appeared from behind the occluder on the Test trials (total 1400 ms, from 8400 ms to 9800 ms into the trial). Time periods included an additional 200 ms to allow for oculomotor planning (McMurray & Aslin, 2004). Time Window 3 began at the end of Time Window 2 and continued until the end of the trial (total 1700 ms, from 9800 until 11,500 ms). Each trial lasted 11,500 ms. The time course of trials is presented in Figure 2.

**Areas of analysis**

The visual display was divided into three areas. The Top Area consisted of the portion of the screen above the occluder. The Left and Right areas corresponded to the areas to the left and right of a vertical line coinciding with the center of the occluder (Figure 3), separated by a buffer zone corresponding to 30 pixels. This buffer zone minimizes any potential errors in gaze estimation up to one degree gaze angle.

---

**Figure 2** Time course of Pre-training, Training and Test trials. During Pre-training trials (n = 4), objects moved in front of the occluder. During Training trials (n = 16), objects moved behind the occluder and reappeared. During Test trials (n = 8), objects moved behind the occluder but did not reappear.
Timing and area of analysis

During Time Window 1, analyses focused on participants’ looking within the Top Area, when the objects moved from side to side above the occluder, and multiple tokens of the word forms were played. Time Window 1 analyses were used to confirm that participants were attending to the word–object pairing. Training and Test trials were not included in further analyses if participants did not attend more than 1.5 s during Time Window 1 to the Top Area. It was essential that participants attended to the specific word–object pairing presented during each trial; otherwise, it would not be possible to interpret participants’ looking behavior after the object went behind the occluder. The 1.5 s window was the minimum time required for infants to be exposed to one auditory token of the word form presented during the trial.

For Training trials, analyses were based on Time Window 2, and focused on the Right and Left areas to determine whether participants made anticipatory eye movements to the correct side of the occluder, i.e. the side on which the object would emerge from behind the occluder. These analyses established whether during training participants had learned that the two pairs of stimuli were associated with different sides of the visual display.

Recall that in the Test trials, word–object pairings were reversed. Therefore, for any one word–object pair during the Test phase, one side of the visual display had been associated with the word form while the other side had been associated with the object (based on the Training trials). For example, if during training, lif–crown consistently emerged from the left side of the occluder and neem–molecule emerged from the right side of the occluder, the pairing of neem–crown presented during the Test phase would be composed of the word form neem (associated during training with the right side) and the object crown (associated during training with the left side). Looking times for Test trials were analyzed during Time Window 3; during this time the object did not reappear from behind the occluder and the word form was not heard. Thus, Test trials could be analyzed for whether participants tracked the word form, by looking to the side of the occluder associated with the word form (in this example, the right side), or tracked the object, by looking to the side of the occluder associated with the object (in this example, the left side).

Results

Results: Training phase

The Training trials were analyzed to establish whether participants had learned that the two pairs of stimuli were associated with different sides of the visual display. We predicted that if participants had learned the pairings, they would spend more time looking to the correct side of the occluder, that is, the side where the object would emerge from behind the occluder. We examined whether learning had occurred over the 16 Training trials. Recall that Training trials were preceded by four Pre-training trials in which the object moved in front of the occluder. Trials were analyzed only if participants had attended for more than 1.5 s during Time Window 1 to the Top Area. Based on this criterion, 11 trials (4.3%) were excluded in the analysis for 18-month-olds, and three trials (1.2%) were excluded in the analysis for the adults.

Analyses were based on the proportion of anticipated looking time to the correct side during Time Window 2, calculated as the proportion of looking time to the correct side as a function of the combined looks to either side. T-tests were used to compare the proportion of anticipated looking time to the correct side compared to chance (0.5) for each group (18-month-olds and adults). Both groups had a higher proportion of anticipated looking time to the correct side, 18-month-olds ($t(15) = 3.09, p < .01$) and adults ($t(15) = 10.17, p < .001$). This indicates that participants looked longer to the correct side (where the word–object pairing was going to reappear) than expected by chance (mean 18 months
Correct: .58, mean Adults Correct: .83). This demonstrates that participants had learned to pair the word–object with the correct side over the course of the Training trials. To assess whether there was a difference between the two groups, an independent sample t-test was used based on the proportion of anticipated looking to the correct side. Eighteen-month-olds had less correct anticipatory looking compared to adults (t(30) = 5.88, \( p < .001 \)). The effect of Anticipation was significant for both the 18-month-olds and for adults, but the adults looked longer to the correct side than the 18-month-olds. The current analyses were based on accumulated looking time to either the correct or incorrect side. However, it is also possible to analyze the data using a categorical analysis, and to examine the proportion of correctly anticipated trials (McMurray & Aslin, 2004). As in the study done by Albareda-Castellot et al. (2011), we found the same results when the analyses were based on accumulated looking time and when they were based on the proportion of correctly anticipated trials.\(^1\)

Results: Test phase

As with the Training trials, the Test trials were analyzed to determine whether participants had attended for more than 1.5 s during Time Window 1 to the Top Area. Based on this criterion, the following trials were excluded: 18-month-olds, First Block Test trials (n = 4, 6.25%); Last Block Test trials (n = 15, 23.4%); Adults, Last Block Test trials (n = 1, 1.5%). Analyses were then conducted for measures collected during Time Window 3.

A 2 \( \times \) 2 repeated-measures ANOVA was used based on the proportion of looking time to the word form, with Block (First, Last) as the within-subjects factor, and Age (18 months, adults) as the between-subjects factor. Block analyses were included to determine whether participants’ looking times varied over the course of the Test trials. There was a significant main effect of Age (\( F(1, 30) = 8.09, \ p < .01, \eta^2 = .21 \)) and a significant two-way interaction between Age and Block (\( F(1, 30) = 5.51, \ p < .05, \eta^2 = .16 \)). Results indicated that looking times to word forms differed for the two groups, and differed between the first and second block of Test trials (Figure 4). To assess this interaction, t-tests with each group compared the proportion of looking time to the word form compared to chance (0.5), separated by Blocks 1 and 2. In Block 1, 18-month-olds had a significantly higher proportion of looking to the word form compared to chance (t(15) = 3.98, \( p < .01, M = .65 \)), suggesting that they relied more on the word forms than on the objects during this task. The adults showed the opposite pattern, relying more on the objects than on the word forms, as indicated by a significantly lower looking to the word form compared to chance (t(15) = 2.53, \( p < .05, M = .38 \)). In Block 2, there were no significant effects and neither group had significantly different looking times to the side associated with the word form compared to chance. The same pattern of results was found when analyzing the data based on the proportion of correctly anticipated trials.\(^2\)

Discussion

This research investigated the underpinnings for two ways in which language learners might begin to make mappings between word forms and objects. One possibility is that learners track salient objects, which, in turn, guides their uptake of labels for those objects. The second possibility is that learners recognize and store word forms, which focuses their attention on the objects that those word forms could be used to refer to. The aim of the current study was to explore the balance between infants’ use of word forms and their referents in processing early word–object associations. Results from the Training trials indicated that over the course of training, both infants and adults learned that the two pairs of stimuli were associated with different sides of the visual display. Results based on analyses of Block 1 of the Test trials showed that infants tended to track the

\(^{1}\) To analyze the proportion of correctly anticipated trials, t-tests compared the proportion of correctly anticipated trials to chance (0.5) for each group (18-month-olds and adults). ‘Correctly anticipated’ was defined as longer looking time to the correct side compared to the other side. Both groups had a higher proportion of correctly anticipated trials, 18-month-olds (t(15) = 2.44, \( p < .05 \)) and adults (t(15) = 11.6, \( p < .001 \)). To assess whether there was a difference between the two groups, an independent sample t-test was used based on the proportion of correctly anticipated trials. Eighteen-month-olds had fewer correct anticipations compared to adults (18-months \( M = .59 \), Adults \( M = .86, t(30) = 5.83, p < .001 \)).

\(^{2}\) A repeated-measures ANOVA was used based on the proportion of trials with greater looking to the word form, with Block (First, Last) as the within-subjects factor, and Age (18 months, adults) as the between-subjects factor. There was a significant main effect of Age (\( F(1, 30) = 11.42, \ p < .01, \eta^2 = .28 \)) and a significant two-way interaction between Age and Block (\( F(1, 30) = 5.20, \ p < .05, \eta^2 = .15 \)). To assess this interaction, t-tests with each group compared the proportion of trials with greater looking to the word form compared to chance (0.5), separated by Blocks 1 and 2. In Block 1, 18-month-olds had a significantly higher proportion of trials looking to the word form compared to chance (t(15) = 2.96, \( p < .01, M = .68 \)). The adults showed the opposite pattern, with a significantly lower proportion of trials looking to the word form compared to chance (t(15) = 3.90, \( p < .01, M = .30 \)). In Block 2, there were no significant effects.
word forms, and that adults demonstrated the opposite tendency, that is, for tracking the objects. The results from Block 2 are not as reliable as those from Block 1. Infants and adults may have learned over the course of the test trials that the object no longer emerged from behind the occluder, leading to a lack of interest, and thus no effect, in Block 2. Further, the extended exposure to the reversed word–object pairings in Block 1 may also have led to inconclusive performance in Block 2.

We have shown that in a word–object association task, 18-month-old infants are guided more by the word form than the object. But we still have not addressed what might account for this early advantage for word forms. We suggest that an advantage for word forms reflects infants’ primitive push towards reference. If word forms are prioritized as the anchors that allow reference to objects, this expectation could aid early word learning (Martin, Onishi & Vouloumanos, 2012; Waxman, 2008; Waxman & Gelman, 2009), and confer special status to word forms as infants build the early word–object associations of their nascent lexicons. Recent results supporting this view come from studies such as those by Ferry et al. (2010), who found that infants at 3 and 4 months generalized categorization of trained exemplars to novel exemplars when exposed to word forms, but not tones. Similarly, Mackenzie, Graham, and Curtin (2010) found that linguistic forms were privileged in their study with 12-month-old infants. Using a Switch task paradigm, they showed that infants formed associations between word forms and objects (fep, wug), but not between communicative sounds (mmm, aaah) and objects, or between single consonant sounds (/l/, /ʃ/) and objects. These findings suggest that the greater reliance on the word forms in word–object association shown by the 18-month-old infants in our work may reflect the privileged status of linguistic stimuli as a foundation for reference.

Our results differ from those of Sloutsky and Robinson (2008) for cross-modal perception. In their studies, 10-month-old infants showed attenuation of the processing of visual input when accompanied by word forms, but 16-month-old infants showed no such auditory dominance effect for linguistic stimuli. Sloutsky and Robinson measured infants’ looking responses to either a brand-new auditory or a brand-new visual stimulus in a perceptual task; thus their work does not shed light directly on how infants weigh the two components with respect to one another in the task. Furthermore, because their visual stimuli consisted of configurations of three geometric shapes, young infants were not likely to approach the task as one of associating the word form with an ‘object’. On the other hand, the current study, using the AEM paradigm, which does allow for a direct comparison of how infants process word forms and objects, found dominance of the word forms over the object stimulus at 18 months of age in a word–object association task. Sloutsky and Robinson suggest that, as infants acquire experience with language, linguistic labels become more familiar, interfering less with the processing of simultaneously presented visual stimuli. Our results demonstrate, however, that this shift does not result in equal attention to word forms and object stimuli in a word–object association task at 18 months. Rather, infants at this age appear to favor word forms in the processing of the cross-modal stimuli in this task.

It might be suggested that the variability present in the auditory stimuli (recall that infants were presented with six different tokens for each of the two words) led to better encoding of the auditory stimuli than of the less variable visual stimuli. This might explain the infants’ reliance on word forms. Previous findings in the literature have shown that stimulus variability can be a factor in perceptual learning (Richtsmeier, Gerken, Goffman & Hogan, 2009; Rost & McMurray, 2009). However, similar auditory effects have been found in studies in which the auditory stimuli did not vary (Sloutsky & Robinson, 2008), suggesting that variability alone will not account for the difference in the current study. Moreover the amount of variability in the acoustic tokens used in the current study was much less than that of Richtsmeier et al. and of Rost and McMurray. There, multiple voices were used, whereas in our case a single voice was used, all stimuli were of the same duration, and only intonation differences distinguished the instances.

Figure 4  Proportion of looking to side associated with word form during Training. Looking times collapsed across both Test word–object pairs, and broken down by age groups and blocks. Error bars indicated standard error.
While we think it is unlikely that the effect reported here was driven by acoustic variability, it might still be the case that the reliance on word forms in word–object association at this age and the importance of the word form in anchoring early reference are rooted in a bias toward the auditory modality in younger infants (also see Lewkowicz, 1988a, 1988b for cross-modal processing). Word–object association in the AEM paradigm is a challenging task. Infants must not only link word forms and objects, but also learn which side of the occluder each pair is associated with. Further, their performance in the task is measured not simply by looking time, but by predictive looking to the correct side. It might be the case that infants respond to the greater demands of this task by falling back, as it were, on the reliance on auditory stimuli they exhibited in purely perceptual tasks at a younger age. The activity of word–object association in the real world is also a demanding task for 18-month-old infants. As in perception, infants must juggle (at least) both auditory and visual stimuli, while also considering the relationship between the two. The reliance on auditory stimuli shown by younger infants in the perceptual domain could manifest in older infants in their reliance on the auditory word form to drive the associations of word and object that support the building of the early lexicon. One would expect then that this reliance or weighing would change depending on the difficulty of the task or amount of exposure to the stimuli. For example, with more exposure to a word–object pairing, one would predict that the relative weighing of the word and object would be more equal compared to a shorter amount of exposure.

The AEM paradigm provides us with a novel way to examine in detail for the first time how word forms and objects each contribute to the formation of early word–object associations. The extent to which the word-over-object weighing is rooted exclusively in a conceptual preparation for reference versus a perceptual bias for the auditory form that subsequently gives way to selective attention to word forms is an exciting and rich area for future study. Relevant research is needed to examine, for example, whether infants at a younger age show the same tendency toward processing word forms as that shown by the 18-month-old infants in this study. Moreover, recall that the adults in our study relied on the object stimuli to process the association; further work within this paradigm is needed to explore the age at which this crossover develops. Other future studies may reveal whether the same reliance on word forms is found in the case of word–object associations with minimally distinct word forms (bin–din). Perhaps learners rely on the word forms to a greater extent when the acoustic dimension is less salient, or perhaps such minimal contrasts focus even greater attention on word forms. Further, an investigation of the relative weighing of non-linguistic sounds (tones, for example) and objects in a similar AEM paradigm would help to reveal to what extent this phenomenon is restricted to the linguistic domain.

In summary, our findings support the claim that infants weigh the word form more heavily than the object, laying the foundation for an approach to word learning in which they ‘may store a sound pattern first and then look to link it to the appropriate meaning’ (Jusczyk, 1997, p. 133). Our work gives a first insight into how infants integrate the complex interplay of knowledge and experience of word forms and objects, as they begin to build a lexicon that incorporates both linguistic and object-relevant information.

**Acknowledgements**

This research was supported by Netherlands Organisation for Scientific Research (NWO) 275-75-001 awarded to TZ, SSHRC 28105 grant awarded to LF, NSERC 81103 awarded to JFW, and a McDonnell Foundation grant (412783-001G) to Richard N. Aslin, PI, JFW, Co-I.

**References**


Received: 12 April 2012
Accepted: 26 September 2013