Word learning in the absence of a speaker

Jason Scofield, Amie Williams and Douglas A. Behrend

First Language 2007; 27; 297
DOI: 10.1177/0142723707077252

ARTICLE

FIRST LANGUAGE

www.sagepublications.com Vol 27(3): 297–311 (200708)
DOI: 10.1177/0142723707077252

Word learning in the absence of a speaker

Jason Scofield, *University of Alabama*
Amie Williams, *University of Alabama*
Douglas A. Behrend, *University of Arkansas*

ABSTRACT
Past studies show that a referential context is important for successful word learning. Still in question, however, is the success of word learning in the absence of a referential context. In this study 2-year-olds were presented with novel words in the absence of a speaker and therefore in the absence of a referential context. Findings revealed that word learning was successful across the experimental trials at rates greater than chance and at rates greater than in the control trial. Findings demonstrated that the absence of a speaker and the referential context provided by a speaker did not result in unsuccessful word learning. It is concluded that a referential context is not necessary for successful word learning.

KEYWORDS
Children; language development; referential context; referential intent; word learning

Word learning is often characterized as occurring during a dyadic interaction between a speaker and a word learner (Clark & Clark, 1977; Naigles & Mayeux, 2001; Rice & Woodsmall, 1988). This interaction generally provides a rich referential context for the word learner, a context that is principally supplied by the presence and actions of the speaker (Baldwin, 2000; Baldwin & Moses, 2001). As a result, it has been suggested that successful word learning depends largely on the word learner's ability to interpret the speaker as intentional and the speaker's role in the interaction as referential (Akhtar, 2005a; Baldwin, 1993a; Baldwin, 1995; Baldwin et al., 1996;

Baldwin & Moses, 2001; Bruner, 1999; Moore, Angelopoulos & Bennett, 1999; Sabbagh & Baldwin, 2005; Tomasello, 1995, 2001).

Successful word learning can be supported by a variety of cues in a variety of contexts (Bloom, 2000). These include, among others: conceptual (Waxman & Markow, 1995), linguistic (Hall, 1994; Hall & Graham, 1999), emotional (Baldwin & Moses, 2001; Tomasello & Barton, 1994; Tomasello, Strosberg & Akhtar, 1996), and referential (Baldwin, 2000; Baldwin & Moses, 2001) cues. Cues such as these are commonly supplied by the speaker and, under ideal learning conditions, are likely to converge on a single target referent. Referential cues are particularly important because they often co-occur with cues from other contexts and, perhaps more significantly, because they often convey information about a speaker's underlying referential intent (Akhtar & Tomasello, 2000; Baldwin, 1995, 2000; Baldwin & Moses, 2001; Tomasello, 1995). Common examples of referential cues include gaze direction, head-turning, voice direction, pointing and joint attention (Baldwin & Moses, 2001).

Using these cues to infer the intentional nature of a speaker's references can help resolve two serious challenges confronting word learners (for discussion, see Scofield & Behrend, 2007). First, inferring referential intent allows the word learner to differentiate between referential acts and non-referential acts (Baldwin, 1993a; Baldwin et al., 1996; Moore et al., 1999). Second, inferring referential intent allows the word learner to differentiate between the referential target and other potential targets (Baldwin, 1993a; Baldwin & Tomasello, 1998; Behrend, 1990; Markman, 1990; Quine, 1960). By conveying both the location of a referent and the speaker's intent to refer to that referent, referential cues can help to resolve these challenges (Baldwin, 1993a; Baldwin et al., 1996; Baldwin & Moses, 2001; Moore et al., 1999). Consequently, it has been suggested that referential cues are important or even, as Baldwin & Moses (2001) described, 'criterial' for successful word learning (see also Akhtar & Tomasello, 2000; Baldwin, 1995; Scaife & Bruner, 1975).

In support of the notion that referential cues are important, there is substantial evidence that word learning is more likely to be successful when referential cues are present. The speaker's role in the delivery of these cues cannot be overestimated. The frequency of mother/child joint attention in infancy has been found to be related to later vocabulary size (Carpenter, Nagell & Tomasello, 1998; Tomasello & Todd, 1983). Through eye gaze, voice direction and other orienting cues provided during a word-learning episode, the speaker also regularly directs (i.e., attention switching) or follows (i.e., attention following or follow-in labeling) the word learner's attention to an intended target (Akhtar, Dunham & Dunham, 1991; Dunham, Dunham & Curwin, 1993; Tomasello & Farrar, 1986).

The word learner's role in recognizing and responding to these cues also cannot be overestimated (Baldwin, 1991, 1993a, 1993b). Baldwin (1991, 1993a) found that the word learner referenced the speaker's gaze, and thereby avoided a mapping error, when the word learner's focus and the speaker's focus did not initially match, (i.e., a 'discrepant labeling' condition); see also Sabbagh & Baldwin, 2005. Similarly, Baldwin and colleagues found that the word learner was more likely to reference the speaker when a word was presented in the presence of two objects rather than one object (Baldwin, 2000). Additional evidence suggests that referential cues, when available, are favored over both the temporal contiguity between the word

and referent and the salience of the referent (Baldwin, 1993a, 2000; Baldwin et al., 1996; Moore et al., 1999).

Interestingly, while the most common referential cues are likely to be overt and inferred directly through the speaker's attention or orientation (Baldwin, 1993a; Hollich, Hirsh-Pasek & Golinkoff, 2000; Moore et al., 1999; Tomasello, 1999), others are inferred indirectly through the speaker's behaviors, emotions and knowledge state. The speaker's behaviors can provide important cues for directing attention towards a particular referent. Tomasello & Barton (1994) found that the word learner learned a novel verb only when the speaker performed an intentional, rather than accidental, action on the target and learned a novel noun only when the speaker intentionally chose, rather than rejected, the target (see also Akhtar, Carpenter & Tomasello, 1996; Akhtar & Tomasello, 1996; Behrend & Scofield, 2006; Diesendruck, Markson, Akhtar & Reudor, 2004; Tomasello et al., 1996). The speaker's emotions can provide important cues for directing attention towards a particular referent. Moses, Baldwin, Rosicky & Tidball (2001) found that, when accompanied by other referential cues, emotions like disgust or pleasure helped to guide the word learner towards the appropriate, intended referent (also see Baldwin & Moses, 2001; Tomasello et al., 1996). Finally, the speaker's knowledge can also provide important cues for directing attention towards a particular referent. Koenig, Clement & Harris (2004) found that the word learner was more likely to link novel words and referents when the speaker reliably, rather than unreliably, labeled objects for which the word learner already had names (also see Diesendruck et al., 2004; Koenig & Echols, 2003; Koenig & Harris, 2005; Sabbagh & Baldwin, 2001). Together, these studies demonstrate that a wide variety of referential cues contribute to creating a rich referential context. Furthermore, this referential context is regularly available during word learning, and the word learner is skilled at using it (Baldwin, 2000; Carpenter et al., 1998).

Indeed, in these studies successful word learning seems to depend exactly on the availability of a supporting referential context and the presence and actions of the speaker who provides it. As a result, it remains largely unclear whether word learning would be successful in conditions in which the speaker and/or the referential context were not present. There is however evidence to suggest that it would. Akhtar, Jipson & Callanan (2001) found that word learning succeeded when the word learner simply overheard, but did not participate directly in, an interaction involving the speaker (see also Akhtar, 2005b; Jaswal & Markman, 2003). While it is possible that the word learner in these studies received indirect support from observing the nearby referential context, the results of Akhtar et al. seem to suggest that the word learner did not need to play an active, central role in that context. Taking these results a step further, Scofield & Behrend (2007) found that word learning succeeded when the word learner was not able to receive referential cues, namely joint attention, to help to establish and maintain a referential context. Interestingly, word learning in both the Akhtar et al. (2001) and Scofield & Behrend (2007) studies succeeded at an impressive rate despite minimal support from the referential context.

Additional evidence for successful word learning in a minimal referential context can be found in classic word-learning studies. In their seminal study, Carey & Bartlett (1978) found that word learning succeeded when a speaker's cues were referentially ambiguous. In that study, a word learner was positioned opposite two objects,

one of a known color and one of an unknown color. The word learner was then asked to select the referent that best corresponded to a novel word (e.g., 'Can you get the chromium one?'). Findings consistently revealed that the word learner selected the object of the unknown color in response to the novel word. These findings clearly demonstrated successful word learning despite the absence of referential cues that could directly differentiate between the target and a distracter (Markman, 1990; Markman & Wachtel, 1988; Merriman & Bowman, 1989; see also Clark, 1990; Diesendruck & Markson, 2001).

One potential limitation of these previous studies is that a speaker was always physically present. As such, it is possible to argue that the mere presence of a speaker created a social context that was sufficient to initiate an adequate, if not an ideal, referential context. If this is true, then perhaps a stronger demonstration of word learning in the absence of a referential context would occur in conditions in which a speaker was physically, rather than just referentially, absent (Grela, Krcmar & Lin, 2004; Rice & Woodsmall, 1988). Other previous studies have actually examined similar conditions. Rice and Woodsmall presented 3- and 5-year-olds with animated videos in which a narrator directly or indirectly presented novel words while reciting a story. During direct presentations the narrator delivered a first-person account of an animated character that oriented to, and labeled, a target, e.g., 'Now, I can't play this on my gramophone.' During indirect presentations the narrator delivered a third-person account of an animated character that oriented to, and labeled, a target, e.g., 'He takes his viola and trudges down the road.' Rice and Woodsmall found that word learning was successful, especially for nouns, in both conditions even though a speaker was not physically present.

Similarly, Grela et al. (2004) presented 15- to 24-month-olds with novel words in either a maximal word-learning condition or minimal word-learning condition. The maximal word-learning condition required unambiguous verbal and nonverbal cues, such as when a novel word was presented by a speaker during a joint reference condition. During the maximal condition, a speaker selected a target from an array of objects, explicitly showed the target to the word learner, and labeled the target five times with a novel word, e.g., 'Here is a sas.' Not surprisingly, the maximal condition resulted in a high rate of word learning, e.g., approximately 85% of trials for older word learners (see also Study 1 of Behrend, Scofield & Kleinknecht, 2001). In contrast, minimal conditions did not require unambiguous verbal and nonverbal cues, such as when a novel word was presented by a speaker during a discrepant labeling condition, by a speaker on a television video, and by a narrator during a segment of *Teletubbies*. Interestingly, Grela et al. found that word learning occurred in all three conditions, even in the *Teletubbies* condition where a speaker was not physically present.

While the results of Rice & Woodsmall (1988) and Grela et al. (2004) suggest that word learning can occur when a speaker is not physically present (see also Naigles & Mayeux, 2001), it is still possible to argue that a referential context was available in these studies. That is, conditions in both studies included either a video of a speaker or a video of a speaker with human-like qualities (e.g., animated mole, bug or periscope). If at any time the word learner believed that these characters

were acting referentially then, technically speaking, the word learner could have acted on a referential context. Perhaps an even stronger demonstration of word learning in the absence of a referential context would occur in conditions that have no dependence on a speaker (i.e., human or otherwise) and therefore provide no access to a referential context.

To date, very few studies have presented a word learner with conditions this referentially stark. Werker, Cohen, Lloyd, Casasola & Stager (1998) presented 14-month-olds with a short video display during which a novel referent moved across the computer screen and an audio speaker presented a novel word. Using a looking-time measure, Werker et al. found that 14-month-olds reliably associated the correct word and referent despite having had no referential context to support the association. Using a conceptually similar procedure and measure, Schafer & Plunkett (1998) also found that an association between a novel word and referent could be formed under conditions in which neither a speaker nor a referential context was available to the word learner. Together, these results have important implications for the success of word learning in the absence of a speaker and the referential context that a speaker provides. However, it is difficult to draw definitive conclusions from the results, in part because looking time is not a robust word-learning measure (Werker et al., 1998).

The studies thus far reviewed provide the grounds for a potentially serious challenge to the notion that a referential context is important for successful word learning. Their clear implication is that the referential context plays a less crucial role than is commonly suggested (Hoff & Naigles, 2002). The aim of the current study was to continue and to expand on the line of inquiry in this research by studying the success of word learning in the absence of a referential context by removing the speaker entirely from the interaction. In the current study the word learner was presented with novel words and novel referents on an animated video. The words were presented via an audio track of a human speaker's voice dubbed directly onto the video. The video and the audio track were important methodological features because they allowed for the presentations to occur in the absence of a speaker. It was hypothesized that if word learning is successful in the absence of a speaker or in the absence of the referential context provided by a speaker, then neither could have been critical (i.e., 'criterial') for successful word learning.

METHOD

Participants

Thirty-eight 2- and 3-year-olds (M = 2;5.7, SD = 0;5.4), 25 girls and 13 boys, initially took part in the study. Eight participants were subsequently excluded for failure to complete the warm-up session or for failure to complete the introductory session. As a result, 30 participants were included in the final sample (M = 2;8.0, SD = 0;5.4), 19 girls and 11 boys. Participants were recruited with parental consent and participant assent from local preschools. All participants received a sticker as a reward for participation.

Materials

Participants completed a warm-up session and watched seven short animated videos (i.e., 30 seconds each), two during an introductory session and five during a testing session. Videos were created using Macromedia Flash animation software. Videos were displayed on a laptop computer that was placed in front of the participant and between the participant and experimenter. The experimenter could not view the computer screen and thus could not watch the videos. Likewise the experimenter did not provide any supporting cues, referential or otherwise, to the participant.

During each video the image of a target object appeared in the center of the screen and was labeled with a word by a pre-recorded audio track dubbed directly onto the video. Importantly, the audio track was of a human speaker's voice but was never the experimenter's voice. The target image disappeared and a selection set of four images appeared in its place, one in each corner of the screen. The four images that comprised the selection set included the original target image and three distracter images (see Fig. 1). Participants were asked, again via a pre-recorded audio track, to select the image that best corresponded to a word. Introductory videos used known targets, distracters, and words. Testing videos used novel targets, novel distracters, and either novel words or neutral comments.

Procedure

All participants were initially presented with a warm-up session and an introductory session. Both sessions were conducted with the participant and the experimenter seated opposite one another at a small table outside the participant's classroom. During the warm-up session the participant was presented with a set of wooden blocks and asked both to manipulate the blocks (e.g., 'Can you build a tower?') and to categorize the blocks along various dimensions (e.g., 'Can you find the blue blocks?'). The warm-up session was designed to allow the participant to acclimatize to the experimenter and the setting and was usually followed immediately by the introductory session.

During the introductory session the participant viewed two animated videos. The video displayed the image of a known target object (e.g., a dog) and presented a known word three times (e.g., 'This is a dog. A dog. It's a dog.') via a dubbed audio track. After the third presentation of the word, the target image disappeared and then reappeared with three additional images of known distracters (e.g., ball, tree and car). The participant was then asked to select the image (i.e., either the target image or a distracter image) that best corresponded to the known word (e.g., 'Can you help me find the dog?'). The introductory session was designed to allow the participant to acclimatize to the upcoming testing session which, for most participants, followed immediately. Those participants ($N = 8$) who did not meet the inclusion criteria on the warm-up session and the introductory session (i.e., were unable to complete the warm-up session due to disinterest, fussiness or distractibility, or did not make the appropriate selections during the introductory videos) were given their choice of a sticker and returned to their classrooms.

During the testing session the participant viewed five animated videos, four experimental and one control. The laptop remained positioned directly in front of the

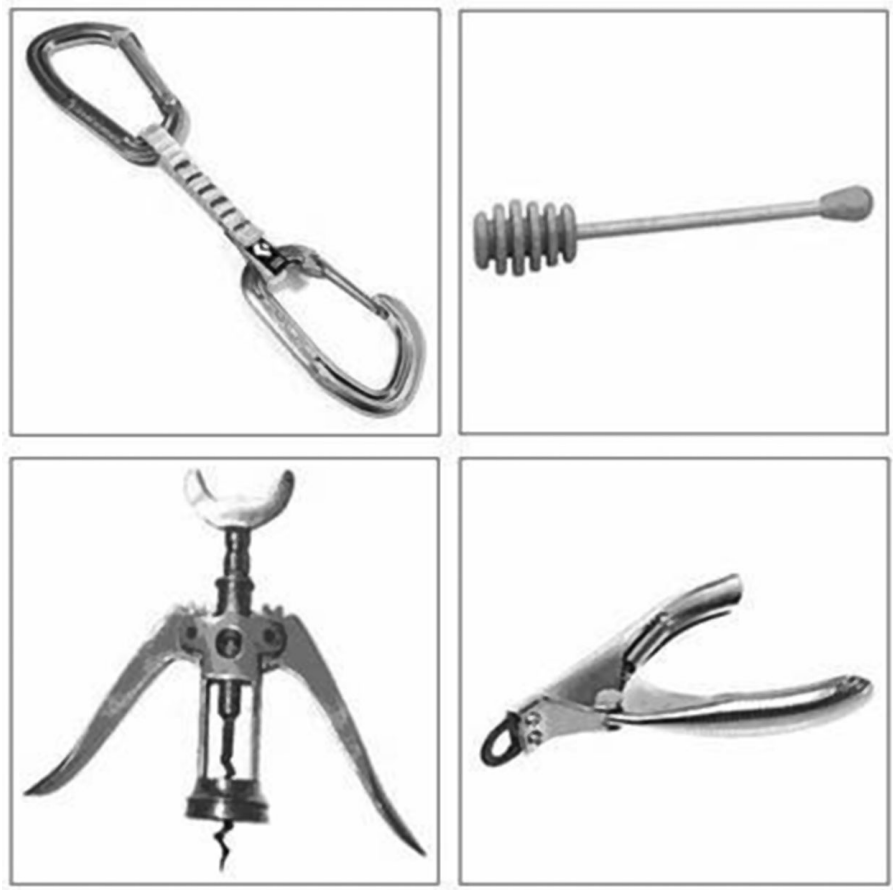

Figure 1 Selection set including a target image and 3 distracter images

participant so that *only* the participant could view the display on screen. During each experimental video, a lone novel target image was displayed on screen and a novel word was presented three times (e.g., 'This is a koba. A koba. It's a koba.'). As in the introductory videos, this word was presented via a dubbed audio track. After the third presentation of the word, the target image disappeared and then reappeared with three additional novel distracter images, giving a total of four images (see Fig. 1). The participant was then asked to select the image that best corresponded to the novel word (e.g., 'Can you help me find the koba?'). The same procedure, using a different novel word, target image and distracter image, was used for the three remaining experimental videos.

The control video was identical to the experimental videos except that a neutral comment, rather than a novel word, was presented three times with the target image (e.g., 'Wow. Neat. Wow.'). The participant was then again asked to select the image that best corresponded to the novel word (e.g., 'Can you help me find the blicket?').

The control video was used to assess the effect of additional exposure to the target image on image (i.e., referent) selection.

After viewing all five videos, the participant was given the choice of a sticker and returned to their classroom. The entire procedure (i.e., warm-up, introductory and testing sessions) lasted approximately 10 minutes. All presentation orders, videos, referents, distracters and words were counterbalanced within each session.

RESULTS

To recap, five trials were completed in the experimental session: four experimental trials and one control trial. In each trial the image of a novel target object was presented on a computer screen with either a novel word (i.e., on the experimental trials) or a neutral comment (i.e., on the control trial). Following each presentation the target image disappeared and then reappeared on the screen (in a novel position) with three additional novel distracter images for a total of four images. The dependent measure on each trial was image selection, either the target or a distracter. The central question was whether the rate of selecting the target on the experimental trials was greater than chance and greater than the rate of selecting the target on the control trial. To answer this question five separate comparisons were conducted: (1) target selection on all experimental trials was compared to chance; (2) target selection on the control trial was compared to chance; (3) target selection on all experimental trials was compared to target selection on the control trial; (4) target selection on each experimental trial was compared to chance; and (5) target selection on each experimental trial was compared to target selection on the control trial.

First, target selection on all experimental trials was compared to chance. In these trials, chance was equal to 25%, as the target was one of four possible selections. If target selection was random, then chance would predict that target selection would occur on 30 of 120 experimental trials. In the current study, target selection occurred on 92 of 120 experimental trials ($M = 0.77$). A one-sample *t*-test revealed that this mean was significantly higher than chance, $t(29) = 9.204$, $p < 0.01$.

Second, target selection on the control trial was compared to chance (i.e., 25%); see Table 1. If target selection was random, then chance would predict that target selection would occur on 7.5 of the 30 control trials. In the current study, target selection occurred on 15 of 30 control trials. Chi-square analysis revealed this proportion was significantly higher than chance, χ^2 (1, $N = 30$) = 10, $p < 0.01$.

Third, target selection on all experimental trials was compared to target selection on the control trial. Target selection in the current study occurred on 92 of 120 experimental trials and on 15 of 30 control trials. As this study used a within-subjects design and the total number of experimental and control trials was not equal, comparing the two sets of trials was difficult. To conduct this comparison the data were transformed to equate the total number of trials in each set, bringing the number of experimental and control trials to 120 for subsequent analysis. Following this transformation, a paired-samples *t*-test revealed that the mean number of target selections on the experimental trials ($M = 0.77$) was significantly higher than the mean number of target selections on the control trial ($M = 0.50$), $t(119) = 5.87$, $p < 0.001$.

Fourth, target selection on each individual experimental trial was compared to chance (i.e., 25%); see Table 1. If target selection was random, then it would occur on 7.5 of 30 trials for each set of experimental trials. In trial 1, target selection occurred on 23 of the 30 trials (77%), χ^2 (1, $N=30$) = 42.711, $p<0.001$. In trial 2, target selection occurred on 22 of the 30 trials (73%), χ^2 (1, $N=30$)=37.378, $p<0.001$. In trial 3, target selection occurred on 27 of the 30 trials (90%), χ^2 (1, $N=30$) = 67.60, $p<0.001$. In trial 4, target selection occurred on 20 of the 30 trials (67%), χ^2 (1, $N=30$) = 27.778, $p<0.001$. On all four experimental trials chi-square analysis revealed that the proportion of target selections was significantly higher than chance.

Fifth, target selection on each individual experimental trial was compared to target selection on the control trial (see Table 1). McNemar analysis revealed that target selection on experimental trial 1 (i.e., 23/30, $p<0.05$), experimental trial 2 (i.e., 22/30, $p<0.05$) and experimental trial 3 (i.e., 27/30, $p<0.001$) were each more likely than target selection on the control trial (i.e., 15/30). McNemar analysis also revealed that target selection on experimental trial 4 (i.e., 20/30) was not more likely than target selection on the control trial.

Overall, these comparisons indicated that: (1) target selection on all experimental trials occurred more often than chance; (2) target selection on the control trial occurred more often than chance; (3) target selection on all experimental trials occurred more often than target selection on the control trial; (4) target selection on each individual experimental trial occurred more often than chance; and (5) target selection on experimental trials 1, 2 and 3 occurred more often than target selection on the control trial. Finally, target selection on experimental trial 4 did not differ from target selection on the control trial.

DISCUSSION

Successful word learning often occurs with support from a referential context, a context that is principally supplied by the presence and actions of a speaker (Baldwin, 2000; Baldwin & Moses, 2001). The current study aimed to examine the success of word learning in conditions in which the speaker, and therefore the referential context,

Table 1 Number and percentage of target selections on experimental, chance and control trials

Trial	*Number*	*Percentage*
Koba	23/30[ab]	77%
Agnew	22/30[ab]	73%
Jeter	27/30[ab]	90%
Nixon	20/30[a]	67%
Chance	7.5/30[b]	25%
Control	15/30[a]	50%

[a] Different from chance, $p<0.05$.
[b] Different from control, $p<0.05$.

was not present. The results indicated that successful word learning (i.e., target selection) occurred on a majority of experimental trials (i.e., 77%) and that the rate of success on these trials differed significantly from the rate on the control trial (i.e., 50%) and from the rate predicted by chance (i.e., 25%). Furthermore, although target exposure did appear to have some effect on target selection, target exposure alone did not explain the difference between the experimental and control trials. Together these results strongly suggest that word learning was uniquely successful in the absence of a speaker and in the absence of a referential context and that this success was not simply due to a slight difference in target exposure. Consequently, these results seriously challenge the notion that a referential context is an important or critical component of successful word learning.

Successful word learning in the current study is really quite extraordinary. The ability to learn a new word after only minimal exposure (i.e., fast mapping) is alone impressive (Carey & Bartlett, 1978). However, in the current study the word was also learned despite a strikingly unsupportive context (Rice & Woodsmall, 1988; Werker et al., 1998). That is, the word learner was unable to physically examine the targets or distracters. Likewise, the word learner was unable both to participate in a social exchange with a speaker and to receive referential cues to highlight the intended referent. Demonstrations of word learning in such stark contexts are rare (see Schafer & Plunkett, 1998; Werker et al., 1998) and have even been characterized as potentially unachievable (Hoff-Ginsberg & Shatz, 1982; Rice, Huston, Truglio & Wright, 1990). During a non-live, non-interactive and non-referential context it would be easy to predict that word learning would fail. After all, successful word learning often occurs in an exact opposite context – live, interactive and referential. However, word learning in the current study did not fail.

The success of word learning during these stark, or minimal, referential conditions is an especially important contribution of this study (Grela et al., 2004; see also Scofield & Behrend, 2007). Although they are among the least supportive, some previous studies have found that minimal conditions actually yield successful word learning at rates similar to those found in other, more supportive conditions (Akhtar et al., 2001; Baldwin, 1991, 1993a; Grela et al., 2004; Rice & Woodsmall, 1988; Scofield & Behrend, 2007). For example, Akhtar et al. presented 2-year-olds with an addressed condition or an overhearing condition. In the addressed condition the word learner and speaker played a finding game in which the speaker retrieved four unfamiliar objects from a box, one of which was labeled three times with a novel word. The overhearing condition was similar to the addressed condition except that the word learner was positioned as an onlooker while the speaker and an assistant completed the finding game. Akhtar et al. (2001) found that a high percentage of word learners (i.e., 83%) in each condition successfully learnt the novel word.

Similarly, Scofield & Behrend (2007) presented 2-year-olds with conditions in which various components of joint attention were absent during labeling. For example, in one condition the word learner's view of the target object was obstructed during labeling and in another condition the speaker's view of the target object was obstructed during labeling. Scofield and Behrend found that a high percentage of word learners (72%) in each condition successfully learnt the novel word. Like the current study, neither the Akhtar et al. (2001) nor the Scofield and Behrend studies provided an optimal

referential context to support word learning. In fact, it would seem that the referential context in the current study provided appreciably less support than in these studies. Whereas a speaker in the other two studies oriented towards and, ultimately, labeled the target, there was no speaker in the current study to occupy this role (also see Schafer & Plunkett, 1998; Werker et al., 1998). And yet successful word learning was not significantly affected by the absence of support from the referential context.

It is this last point that may have important implications for theories that stress the significance of a referential context in successful word learning (Akhtar, 2005a; Baldwin, 1995; Baldwin & Moses, 2001; Bruner, 1999; Tomasello, 1995, 2001). Learning a word generally requires the word learner to identify the presence of the word (Baldwin et al., 1996; Moore et al., 1999) and the location of the target (Baldwin & Tomasello, 1998; Behrend, 1990; Markman, 1990; Quine, 1960). Referential cues are a common source for this information. For example, joint attention, a shared attentional state between the speaker and word learner, assists in both identifying the word and locating the referent (Akhtar, 2005a; Baldwin, 1995; Scofield & Behrend, 2007). During joint attention the word learner infers that joint attention is a referential framework which includes the speaker, the word learner and the target (Akhtar, 2005a; Bakeman & Adamson, 1984; Baldwin, 1995; Baldwin et al., 1996; Baldwin & Moses, 2001; Moore et al., 1999) and that the target is the intended referent (Baldwin, 1991, 1993a, 1995; Baldwin et al., 1996). In this way, the referential context is supplied through referential cues like joint attention which, in turn, support successful word learning.

However, referential cues like joint attention also require the presence of the speaker. In fact, much of the referential context is supplied by the presence and actions of the speaker and there is reason to believe that the word learner understands the importance of the speaker's role in this context. For example, when presented with a novel word the word learner frequently references the speaker, paying particular attention to the speaker's eyes, orientation or other cues that may communicate the speaker's intent (Baldwin, 1991, 1993a, 1993b; see also Sabbagh & Baldwin, 2005). It therefore seems reasonable to expect that, powerful as a referential context and a speaker's role within that context can be, word learning would suffer dramatically in their absence. However, the results of the current study contrast with this expectation.

A primary concern is how best to interpret these results, although this should not be done without first addressing some potential criticisms. One possible criticism of the current study is that the novel words and neutral comments were presented by a human voice; this implies that the human voice alone supplied a referential context that supported word learning. It is unlikely that this criticism would be made because referential cues are generally conveyed by a speaker's eyes and/or actions, e.g., gaze, joint attention, pointing, etc. (Baldwin & Moses, 2001), both of which were unavailable, and because word learning differed across conditions whereas the presence of the human voice did not. In fact, only the presence or absence of a novel word differed across conditions. This strongly suggests that the word, and not the voice, explains word learning. Using a synthesized voice in future studies may help resolve this criticism.

A second, related criticism of the current study is that the novel words and neutral comments occurred in the presence of the experimenter; this implies that the experimenter was assigned a referential, rather than a non-participatory, role. Though

possible, it is also unlikely that this criticism would be made. The experimenter conveyed exactly zero referential cues that would differentiate between the target and a distracter (e.g., no joint attention, no gaze, no orienting, etc.) and yet the word learner consistently differentiated between the two. This suggests that the presence of the experimenter alone does not explain the success of word learning. Physically removing the experimenter in future studies may help resolve this criticism.

Another criticism of the current study is that the novel words were presented in a simplified context, one that did not require referential support. This implies that referential support is necessary only when the learning context is complex, such as when a reference is ambiguous. Here, the learning context was decidedly not complex: there was only one potential referent and it was labeled three times (see also Shafer & Plunkett, 1998; Werker et al., 1998). As such, the importance of the role of referential support in a complex learning context is not addressed. It is quite possible that the importance of referential support correlates positively with the complexity of the learning context. However, this particular criticism would be surprising because it would suggest that the necessity of referential support varies with the complexity of the learning context. Creating a more complex learning context in futures studies may help resolve this criticism.

A final criticism of the current study is that 32-month-olds, the average age of participants, are maturing word learners who do not rely on referential support. This implies that referential support is only critical during the early stages of word learning (Baldwin & Moses, 2001; Hoff & Naigles, 2002; Nazzi & Bertoncini, 2003). Along these lines, Baldwin & Moses (2001) argue that, although a diverse and rich set of referential cues are available within the learning context, it may be beneficial for the word learner to be flexible and utilize all available cues. Future studies could include younger participants in order to resolve this criticism.

CONCLUSIONS

Word learning often occurs with a speaker in a context rich with referential support (Baldwin, 2000; Baldwin & Moses, 2001). It has even been suggested that successful word learning results, in part, from the word learner's ability to understand a speaker's role in this context as both intentional and referential (Akhtar, 2005a; Baldwin, 1995; Baldwin & Moses, 2001; Bruner, 1999; Tomasello, 1995, 2001). The current results seriously challenge this suggestion. In this study word learning was successful in the absence of a speaker and therefore in the absence of a referential context. If a referential context truly occupied a necessary or 'criterial' role in the success of word learning, then word learning should have been unsuccessful in this study. As word learning was not unsuccessful, these results can only be interpreted as evidence against the referential context as a necessary or 'criterial' part of successful word learning.

REFERENCES

Akhtar, N. (2005a). Is joint attention necessary for early word learning? In C. Tamis-LeMonda & B. D. Homer (Eds), *The development of social cognition and communication* (pp. 165–179). Mahwah, NJ: Erlbaum.

Akhtar, N. (2005b). The robustness of learning through overhearing. *Developmental Science, 8*, 199–209.
Akhtar, N., Carpenter, M. & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635–645.
Akhtar, N., Dunham, F. & Dunham, P. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language, 18*, 41–49.
Akhtar, N., Jipson, J. & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72*, 416–430.
Akhtar, N. & Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology, 14*, 79–93.
Akhtar, N. & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff & K. Hirsh-Pasek (Eds), *Becoming a word learner: A debate on lexical acquisition* (pp. 114–135). Oxford: Oxford University Press.
Bakeman, R. & Adamson, L. (1984). Coordinating attention to people and object in mother-infant and peer-infant interactions. *Child Development, 55*, 1278–1289.
Baldwin, D. A. (1991). Infants' contribution to the achievement of joint attention. *Child Development, 63*, 875–890.
Baldwin, D. A. (1993a). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832–834.
Baldwin, D. A. (1993b). Infants' ability to consult the speaker for clues to word meaning. *Journal of Child Language, 20*, 395–418.
Baldwin, D. A. (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: Erlbaum.
Baldwin, D. A. (2000). Interpersonal understanding fuels knowledge acquisition. *Current Directions in Psychological Science, 9*, 40–45.
Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, R. N. & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67*, 3135–3153.
Baldwin, D. A. & Moses, L. M. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 309–329.
Baldwin, D. A. & Tomasello, M. (1998). Word learning: A window on early pragmatic understanding. In E. Clark (Ed.), *The proceedings of the twenty-ninth annual child language research forum* (pp. 3–23). Chicago, IL: Center for the Study of Language and Information.
Behrend, D. A. (1990). Constraints and development: A reply to Nelson, 1988. *Cognitive Development, 5*, 313–330.
Behrend, D. A. & Scofield, J. (2006). Verbs, actions, and intentions. In K. Hirsh-Pasek & R. Golinkoff (Eds), *Early verb learning: Action meets words*. Oxford: Oxford University Press.
Behrend, D. A., Scofield, J. & Kleinknecht, E. E. (2001). Beyond fast mapping: Young children's extensions of novel words and novel facts. *Development Psychology, 37*, 698–705.
Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.
Bruner, J. S. (1999). The intentionality of referring. In P. D. Zelazo, J. W. Astington & D. R. Olson (Eds), *Developing theories of intention: Social understanding and self-control* (pp. 329–339). Mahwah, NJ: Erlbaum.
Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17–29.
Carpenter, M., Nagell, K. & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63* (4, Serial No. 255).
Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*, 417–431.
Clark, H. & Clark, E. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Diesendruck, G. & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630–641.
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.
Dunham, P. J., Dunham, F. & Curwin, A. (1993). Joint-attentional states and lexical acquisition at 18 months. *Developmental Psychology, 29*, 827–831.
Grela, B., Krcmar, M. & Lin, Y., (2004, May). Can television help toddlers acquire new words? Retrieved 17 May 2006 from: http://www.speechpathology.com/articles/article_detail.asp?article_id=72.
Hall, D. G. (1994). Semantic constraints on word learning: Proper names and adjectives. *Child Development, 65*, 1299–1317.
Hall, D. G. & Graham, S. A. (1999). Lexical form class information guides word-to-object mapping in preschoolers. *Child Development, 70,* 78–91.
Hoff, E. & Naigles, L. R. (2002). How children use input to acquire a lexicon. *Child Development, 73*, 418–433.
Hoff-Ginsberg, E. & Shatz, M. (1982). Linguistic input and the child's acquisition of language. *Psychological Bulletin, 92*, 3–26.
Hollich, G., Hirsh-Pasek, K. & Golinkoff, R. M. (2000). Breaking the language barrier: An Emergentist Coalition Model of word learning. *Monographs of the Society for the Research in Child Development, 65* (3, Serial No. 262).
Jaswal, V. K. & Markman, E. M. (2003). The relative strengths of indirect and direct word learning. *Developmental Psychology, 39*, 745–760.
Koenig, M. A., Clement, F. & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.
Koenig, M. A. & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and speakers who use them. *Cognition, 87*, 179–208.
Koenig, M. A. & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14*, 57–78.
Markman, E. M. & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 121–157.
Merriman, W. E. & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54* (3–4, Serial No. 220).
Moore, C., Angelopoulos, M. & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*, 60–68.
Moses, L. J., Baldwin, D. A., Rosicky, J. G. & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development, 72*, 718–735.
Naigles, L. R. & Mayeux, L. (2001). Television as incidental language teacher. In D. G. Singer & J. L. Singer (Eds), *Handbook of children and the media* (pp. 135–152). Thousand Oaks, CA: Sage.
Nazzi, T. & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science, 6*, 136–142.
Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
Rice, M. L., Huston, A. C., Truglio, R. & Wright J. C. (1990). Words from Sesame Street: Learning vocabulary while viewing. *Developmental Psychology, 26*, 421–428.

Rice, M. L. & Woodsmall, L. (1988). Lessons from television: Children's word learning from viewing. *Child Development, 59*, 420–429.

Sabbagh, M. A. & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070.

Sabbagh, M. A. & Baldwin, D. A. (2005). Understanding the role of communicative intentions in word learning. In N. Eilan, C. Hoerl, T. McCormack & J. Roessler (Eds), *Joint attention: Communication and other minds: Issues in philosophy and psychology* (pp. 165–184). New York: Clarendon Press/Oxford Press.

Scaife, M. & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253*, 265–266.

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

Scofield, J. & Behrend D. A. (2007). *Word learning in the absence of joint attention.* Manuscript in preparation, University of Alabama.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds), *Joint attention: Its origins and role in development.* Hillsdale, NJ: Erlbaum.

Tomasello, M. (1999). Having intentions, understanding intentions, and understanding communicative intentions. In P. D. Zelazo, J. W. Astington & D. R. Olson (Eds), *Developing theories of intention: Social understanding and self-control* (pp. 63–75). Mahwah, NJ: Erlbaum.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Tomasello & E. Bates (Eds), *Language development: The essential readings* (pp. 111–128). Malden, MA: Blackwell Publishers, Inc.

Tomasello, M. & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology, 30*, 639–650.

Tomasello, M. & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*, 1454–1463.

Tomasello, M., Strosberg, R. & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language, 23*, 157–176.

Tomasello, M. & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4*, 197–211.

Waxman, A. R. & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257–302.

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

ADDRESS FOR CORRESPONDENCE

Dr Jason Scofield
Department of Human Development and Family Studies,
Box 870160, University of Alabama,
Tuscaloosa, AL 35487-0160, USA
E: scofield@ches.ua.edu