Journal of Memory and Language **46,** 39–56 (2002)
doi:10.1006/jmla.2001.2799, available online at http://www.academicpress.com on

Recognition of Words Referring to Present and Absent Objects by 24-Month-Olds

Daniel Swingley

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

and

Anne Fernald

Stanford University

Three experiments tested young children's efficiency in recognizing words in speech referring to absent objects. Seventy-two 24-month-olds heard sentences containing target words denoting objects that were or were not present in a visual display. Children's eye movements were monitored as they heard the sentences. Three distinct patterns of response were shown. Children hearing a familiar word that was an appropriate label for the currently fixated picture maintained their gaze. Children hearing a familiar word that could not apply to the currently fixated picture rapidly shifted their gaze to the alternative picture, whether that alternative was the named target or not, and then continued to search for an appropriate referent. Finally, children hearing an unfamiliar word shifted their gaze slowly and irregularly. This set of outcomes is interpreted as evidence that by 24 months, rapid activation in word recognition does not depend on the presence of the words' referents. Rather, very young children are capable of quickly and efficiently interpreting words in the absence of visual supporting context.

Key Words: word recognition; language acquisition; child; visual fixation; lexicon.

In the second year of life, children become increasingly skilled language users. This progress is most often characterized in terms of countable "acquisitions": the number of words in the child's vocabulary, the length of his or her utterances, the appearance of grammatical morphemes, and so on. Other crucial developments are more gradual in character and are often relatively difficult to assess. Among these is children's improving ability to understand words in a broad range of situational contexts. One important aspect of this change is that children come to demonstrate understanding of what a word refers to even when the referent is not present. The emergence of this ability was described in one of the first observational studies of a child learning language. In this 1787 work, Dietrich Tiedemann noted, "certain names of very familiar objects [the child] understood perfectly, so that even in their absence he had the image of them in mind and looked around in order to point them out" (p. 222; trans. Murchison & Langer, 1927).

Since then, many researchers have noted that children's first words appear to undergo a process of decontextualization, both in comprehension and production. At first, a word may be used only in a very restricted set of circumstances; to take one well-known example, a child might first use the word "car" only to refer to cars seen through a living-room window (Bloom, 1973, p. 72) before going on to extend the word to other situations. A counterpart to this phenomenon in comprehension is the early understanding of the relation between a word and its referent only when the referent is available in the scene. In both cases, development in-

---

This work was supported by grants from the National Institutes of Health, (MH-41511) to A.F. and (F32-HD08307) to D.S. Portions of the research reported here were presented at the 24th Annual Boston University Conference on Language Development. The authors thank Heather Birks, Tomoko Wakabayashi, and the students at the Center for Infant Studies at Stanford for their careful and dedicated assistance. We also thank the many parents and children who made this work possible through their participation.

Address correspondence and reprint requests to Daniel Swingley, Max-Planck-Institute for Psycholinguistics, Postbus 310, 6500-AH Nijmegen, The Netherlands.

0749-596X/01 $35.00

volves reducing reliance on contextual cues that remind the child of a word's meaning (e.g., Bloom, 1993, Chap. 5). This decontextualization is linked to the searching behavior noted by Tiedemann, in that decontextualized words can evoke the notions that the words represent without those notions having been first "primed" by the child's environment.

The tendency to search for an absent object when the object is named typically emerges between the ages of 12 and 16 months (e.g., Bates, 1979; Bloom, 1973; Huttenlocher, 1974; Lewis, 1937; Taine, 1877). Children's initial capacities in this regard appear to be fragile. For example, in a study of 13-month-olds, Snyder, Bates, and Bretherton (1981) classified more than half of the nouns infants knew as "contextually restricted"; for many nouns in their receptive vocabularies, infants failed to show evidence of understanding the words when the referent was absent. However, in the second half of the second year, children's performance in understanding words that are not exemplified in the environment improves. Sachs and Truswell (1978) found that 16- to 24-month-olds were able to carry out actions such as "Smell the truck" about 60% of the time; here, the objects were available from a visible array, but the actions had not been modeled by adults. Similarly, the child studied by Savage-Rumbaugh et al. (1993, p. 78) from the age of 18 to 24 months was able to carry out complex commands like "Go outdoors and get the potato" on about 50% of trials, a figure which underestimates comprehension because there were presumably cases in which the child understood the request but did not comply.

Studies of word learning provide further evidence of maturation in children's ability to link words and meanings. By the age of 18 months, children can learn the meanings of words in "nonostensive" situations in which the new word is uttered in the absence of the referent (Tomasello, Strosberg, & Akhtar, 1996), and between 18 and 24 months, children become sensitive to various signs of the intentions of speakers labeling objects (Akhtar, Carpenter, & Tomasello, 1996; Baldwin, 1991, 1993a, 1993b; Tomasello & Barton, 1994). These experiments all provide evidence of a strong link between ideas and words even when purely perceptual aspects of the situation, such as spatiotemporal contiguity of the label and its referent, do not support this link.

However, while these results point to a decline in children's dependence on visual contextual cues to word recognition over the second year, rapid and efficient word retrieval may still require contextual support. This proposal has been made by Walley (1987), who argued that the nature of children's speech processing differs according to how constrained the situational context is. In constrained contexts, children may pay particular attention to word-initial phonetic information; however, "when there is no context to suggest the identity of a word, children attend more closely to word-final input" (p. 164; see also Cole & Perfetti, 1980). This suggestion was based upon research using a mispronunciation-detection task in which familiar target words were mispronounced either word-initially or word-finally. Some previous experiments with adult participants have shown that detection of initial mispronunciations is easier than detection of final mispronunciations (e.g., Cole, Jakimik, & Cooper, 1978). This effect might be attributed to the fact that adults recognize words incrementally as they hear them, and therefore do not attend closely to word-final phonetic information because it is not as informative as word-initial information. Diminished attention could lead to reduced accuracy in detecting mispronunciations. On this account, word-position effects in mispronunciation detection reflect the incremental nature of spoken word recognition.

To assess the development of the word-position effect, Walley (1987) tested 4- and 5-year-olds' detection of word-initial and word-final mispronunciations. Children were generally more accurate in detecting word-initial mispronunciations than word-final ones when the target word was either (a) placed at the end of a contextually constraining sentence or (b) presented along with a picture of the target's referent. However, the word-initial detection advantage was *not* found when the target word was presented in isolation, without either contextual cue. This was interpreted as evidence that children only attend more to word-initial information than to word-final information when contextual constraints are available. Because speed and efficiency in word recognition are tied to the ability to attend to word-initial information and to process speech continuously. Walley's proposal implies that young children's recognition of words referring to absent objects may be slow or inefficient.

By contrast, adults' recognition of words is rapid and efficient even when words are isolated from linguistic or situational context, as described by several current models of spoken word recognition (e.g., Marslen-Wilson, 1987; Norris, 1994; McClelland & Elman, 1986). Briefly, these models all hold that word recognition is a continuous process in which the listener updates his or her interpretation of the speech stream as it unfolds. This interpretation proceeds so quickly that several words that are phonologically similar at onset may all be activated as the word is spoken: thus, for example, hearing the *tur* of *turkey* leads adults to briefly consider both "turtle" and "turkey," among other words, as potentially intended by the speaker.[1] Rapid activation on the basis of partial phonetic information contributes to the speed with which adults understand spoken language; listeners need not wait for an utterance or even a word to be complete before interpretation can begin.

In adults, evidence of this rapid semantic activation is found even when contextual support in the form of semantically constraining linguistic or environmental information is absent (e.g., Zwitserlood, 1989). This is not to say that context is irrelevant. Generally speaking, word recognition is faster and more reliable when words are sensibly related to their linguistic or real-world context than when they are not (e.g., Marslen-Wilson & Tyler, 1980; Miller & Isard, 1963; Nittrouer & Boothroyd, 1990), and facilitatory effects of semantic context have been shown in hundreds of studies of language understanding (e.g., Meyer & Schvaneveldt, 1971). The generalization we wish to emphasize is that even without relevant contextual information, adults tend to recognize words rapidly and without difficulty.

Recent research has shown that children between the ages of 18 and 24 months already share adults' ability to interpret speech as it unfolds, at least in situations providing clear visual context (Fernald, Swingley, & Pinto, in press; Swingley, Fernald, & Pinto, 1999). These studies used a visual fixation procedure (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Swingley, Pinto, & Fernald, 1998). Children were shown pairs of pictures on computer monitors. One of these pictures was named, and children's eye movements were monitored. Children tended to rapidly fixate the named picture, providing a measure of speed and accuracy in word recognition (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). In Fernald et al. (in press), 18- and 21-month-olds' recognition of whole words (e.g., *baby*) and partial words (*bay*—) was tested. As the words or fragments were presented, children quickly shifted their gaze to fixate the appropriate pictures. This rapid response was identical in the whole-word and partial-word conditions.

Swingley et al. (1999) demonstrated effects similar to these in both 24-month-olds and adults, using a related method. Children viewed picture pairs such as *dog–tree*, in which the names for the pictures did not overlap at onset, and *dog–doll*, in which the names overlapped. Response latencies to the spoken target *doggie* varied according to whether children were initially fixating the tree (the distracter on baseline trials) or the doll (the distracter on overlap trials). Responses were slower in the overlap condition, because more phonetic information was needed to disambiguate *doggie* from *doll* than from *tree* as *doggie* was heard. The same effects were found in adults, using the same stimuli. These results were interpreted as evidence that 24-month-olds, like adults, process speech continuously.

However, as described previously, adults are also fast and efficient in recognizing words

---

[1] Under some circumstances, words not starting with *tur*, such as *dirty*, may be activated as well; word recognition is not "strictly left to right." Onset mismatches appear to significantly hinder, though not necessarily prevent, word recognition (see Allopenna, Magnuson, & Tanenhaus, 1998; Connine, Blasko, & Titone, 1993; Marslen-Wilson & Zwitserlood, 1989).

when words are spoken out of context. The fact that one-year-olds only gradually develop the ability to understand words at all in the absence of their referents suggests that context might have a much greater effect on word recognition in young two-year-olds than in adults. In fact, as Walley (1987) suggested, even children of four or five years may attend to words in a very different manner when the range of possible words is unconstrained by contextual information.

The purpose of the current research was to assess the speed 24-month-olds' recognition of words referring to present and absent objects. Although previous observational studies have shown that children at this age can understand words denoting absent objects or events, the experiments reported here are the first to evaluate the potential effects of visual context on the *processing* of familiar words.[2] A visual fixation procedure was used. Children's word recognition was tested when the referents of spoken words were absent, and also when the referents were present. In Experiment 1, children heard sentences containing a known word which corresponded to one of the two familiar objects depicted (baseline trials) and sentences containing a known word which did not correspond to either of the two familiar objects depicted (mismatch trials). These mismatch trials provided the crucial test of decontextualization because they measured children's speed and accuracy in recognizing spoken words referring to unpictured objects.

EXPERIMENT 1

Previous studies examining the temporal characteristics of very young children's recognition of words have used visual fixation procedures (Fernald et al., 1998; Schafer & Plunkett, 1998; Swingley et al., 1999; Swingley & Aslin, 2000). In our research using this procedure, we have considered separately those trials on which children happened to be fixating the distracter when the target word began ("D-onset trials") and trials on which children happened to be fixating the target when the target word began ("T-onset trials"). Correct performance requires shifting fixation only on D-onset trials and maintaining fixation on T-onset trials. By 24 months, children are good at this; typically, children shift away from the distracter at least 75% of the time and away from the target only about 25% of the time. Children's latency to initiate a shift away from the distractor picture declines with age; 24-month-olds usually take about 700–800 ms, starting from the onset of the target word (e.g., Swingley, Pinto, & Fernald, 1999). In previous studies of this sort, the target picture was always available for examination before the target word was presented.[3] Consequently, children's responses in these studies are not informative about the speech referring to absent entities.

In Experiment 1, we examined children's responses to familiar words in sentences like "Where's the doggie?" On some trials (baseline trials), the target was pictured; on other trials (mismatch trials), the target was not pictured. If children rapidly shift their fixation both on D-onset baseline trials and on mismatch trials, but do *not* shift on T-onset trials, it would show that (1) the pattern of eye movements is contingent on the match between the picture and the target word and that (2) these eye movement responses do not depend on the presence of a matching picture in the display. This pattern of results would suggest that 24-month-olds are not strongly context-bound in their processing of speech. Alternatively, if children are poor at recognizing words out of context, these rapid responses would not be expected on mismatch trials, because on these trials children would not have

---

[2] Even 8-month-olds recognize words as familiar sound-patterns (e.g., Jusczyk & Hohne, 1997), but recognizing a sound-pattern as familiar is not the same as *understanding* a word, which requires retrieval of semantic content; see, e.g., Fernald, McRoberts, and Swingley (2001).

[3] With one exception: Naigles and Gelman (1995), in a study using preferential looking to evaluate children's overextensions, also included some mismatch trials on which a familiar label was applied to two nonmatching pictures. However, differences in the trial timing, coding methods, and data presentation between that study and the experiments described here were too great to permit detailed comparison of the outcomes, though their results were not inconsistent with those reported here.

had the benefit of concurrent presentation of the target word's referent.

Methods

Participants were 24 24-month-olds. On each trial, a picture of a familiar object was presented to children on each of two horizontally aligned computer monitors. A few seconds later, a prerecorded sentence was played. On "baseline" trials, one of the pictured objects was named (e.g., *Where's the doggie?* in the context of a dog and a ball). On "mismatch" trials, a similar sentence named an object not displayed (e.g., *Where's the doggie?* in the context of a shoe and a ball). On all test trials, the first sentence was followed by a second, uninformative sentence (e.g., *Do you see it?*). Children's visual fixations were recorded and coded off-line by coders who noted the timing of stimulus onsets and changes in children's fixations.

Participants. The mean age of the 24 participants was 106.9 weeks (range 104.9 to 108.7). Half were girls. All children were full-term well-baby births, and all children's caregivers had estimated that at least 80% of children's speech input was in English. An additional two children were tested but were excluded from the final sample because they did not complete at least 15 of the 18 test trials.

Visual stimuli. The visual stimuli were digitized photographs of objects on a gray background, presented on 15-inch Apple color monitors. Pictures on test trials included a baby, a ball, a car, a dog, a duck, and a shoe. Pictures were of similar sizes and had been found to be of roughly equal salience to young children in previous studies.

Auditory stimuli. The speech stimuli were recorded by a female native speaker of American English using a Revox B77 reel-to-reel tape recorder. Her speaking rate was slow and in a moderately "infant-directed" register. The *Where's the* portion of each test sentence averaged 600 ms in length. The duration of each target word (in ms) was as follows: *baby*, 866; *ball*, 808; *car*, 723; *doggie*, 754; *duck*, 664; and *shoe*, 726 (mean, 757). A 1000-ms pause followed the offset of each target word; then an additional sentence began, *Do you see it?* for targets *ball*, *doggie*, and *shoe*, and *Can you find it?* for targets *baby*, *car*, and *duck*. These additional sentences were included to help maintain children's interest in the procedure. Sentences for eight filler trials were also recorded, using two additional target words (*kitty* and *birdie*). All sentences were digitized at 22,050 Hz using AudioMedia software for experimental presentation.

Apparatus and procedure. The experiment was conducted in a three-sided cloth-walled booth measuring 1 m by 1.2 m by 2 m tall. The parent sat on a chair in the open end of the booth, holding her child on her lap facing the two monitors, which formed part of the back wall of the booth. The monitors were separated horizontally by about 60 cm, and the child was positioned about 80 cm from the back wall of the booth. Speech stimuli were delivered through a concealed central speaker beneath the monitors. The child's eye movements were observed using a videocamera placed between and slightly below the monitors. The parent's view of the monitors was completely occluded by a black curtain between the child and parent. The procedure was controlled by an experimenter in an adjacent room.

The parent and child were led into the testing room by a second experimenter. The parent signed a consent form and the experimenter described the procedure while trying to help the child feel at ease. Parents had been asked to complete a Communicative Development Inventory (Words and Sentences: Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994), which was collected by the experimenter.

The parent was then seated on a chair in the booth, with her child on her lap. As the curtain was lowered, two identical pictures of trucks were displayed on the monitors. The first experimenter, speaking through a microphone from the adjacent control room, encouraged the child to look at the trucks. This served to familiarize children with the experience of being spoken to by an unseen person. Once children were attending to the truck pictures, the second experimenter left and the first trial began.

The experiment consisted of 26 trials, including 18 test trials and 8 filler trials. Filler trials were included only to add variety to the presentation; responses on these trials were not analyzed. Of the test trials, 12 were baseline trials and 6 were mismatch trials. Each trial began with a 3-s familiarization period in which the two pictures were presented simultaneously without any accompanying speech. This gave children a chance to look at both pictures before hearing the target word. After the familiarization period, the first of the two sentences began. The trial ended 6 s after the onset of the first sentence. Trials were separated by a 1-s pause, during which the monitors were black.

Four stimulus orders were created, the third and fourth being left/right reflections of the first and second. In each order, each of the six test pictures appeared six times: twice as the target, twice as the distracter, and twice on mismatch trials (on which the target/distracter distinction did not apply). Test pictures were grouped so that the car, baby, and duck were paired equally often, as were the ball, dog, and shoe (e.g., the baby appeared with the car and with the duck three times each). Within each order, each picture served as the target once on the left and once on the right, and appeared on the left and right equally often. Most pictures appeared equally often in the first and second halves of the experiment, and each of the six pictures served as the target once in each half of the experiment. Target side on baseline trials was quasirandomly ordered such that a given side was the target side for at most two consecutive trials. Target words on all mismatch trials had previously been heard as targets on baseline trials. Finally, no picture appeared twice on consecutive trials. The entire procedure took about 5 min.

Coding. During recording, videotapes of the children were time-stamped with a digital stopwatch identifying each video frame (33-ms intervals). This enabled coders to make accurate measurements of looking times to the left and right pictures by examining, frame by frame, each change in the location of children's fixations. Coding was done by several highly trained coders who were unaware of the auditory stimulus or target side on each trial. Coders' judgments were then coordinated with information about target side and the timing of the speech stimulus, using custom software.

Analyses. The same set of analysis procedures was used for all experiments. As in previous research, we established a "window" of time during which fixation responses were examined. This window began 367 ms after the onset of the target word, where the word onset was considered to be the beginning of the stop gap initiating the onset consonant (or the beginning of the frication in "shoe"). Most earlier eye movements cannot plausibly be considered responses to the spoken target word, because the mobilization of an eye movement in infants is generally assumed to require a *minimum* of about 200 ms, with a mean considerably higher (e.g., Haith, Wentworth, & Canfield, 1993; see also Canfield, Smith, Brezsnyak, & Snow, 1997). Similar criteria are used in research using fixations to study word recognition in adults (e.g., Dahan, Swingley, Tanenhaus, & Magnuson, 2000). The window of analysis ended 2000 ms after the onset of the target word.

Within this analysis window, we report three measures: *target fixation proportion*, *response latency*, and *shift proportion*. *Target fixation proportion* is defined as the time children fixated the target divided by the total time children fixated the target and distracter. Target fixation proportion provides a measure of children's overall performance on baseline trials across the window of analysis. This measure is not calculated for mismatch trials, which have no target or distracter. *Response latency* (or RT) is defined as the length of time between the onset of the target word and children's first initiation of a shift from one picture to the other. This measure is calculated only for baseline D-onset trials and mismatch trials. As described above, RTs under 367 ms and over 2000 ms are not counted. Response latency is a standard measure in studies of word recognition. *Shift proportion* is the proportion of trials on which children shifted from the initially fixated picture to the other picture. This measure is calculated only for mismatch trials and for baseline D-onset trials; these sets of trials are comparable because in both cases the fixated picture is not the target. High shift proportions on baseline D-onset trials indicate good performance, because it is an error *not* to shift. Shift proportion is roughly analogous to the inverse of the error rate measure typically computed in more common psycholinguistic tasks such as lexical decision. If fast RTs occur with large error rates (here, low shift proportions), the RTs may not be representative of children's processing speed over the set of tested words.

Results

The time-course of children's eye movements in Experiment 1 is pictured in Fig. 1, providing an overview of the results. Each curve represents a different group of trials. For all three curves, the *y*-axis represents the proportion of trials on which children are at that moment fixating a picture different from the picture they had been fixating at target onset. (By definition, then, at time zero, the curves have *y*-values of zero.) The uppermost curve (unfilled diamonds) displays responses for baseline D-onset trials. Children shifting away from the distracter raise the curve; children then shifting back to the distracter lower the curve (note that the graph does not display cumulative frequency). Children usually (correctly) shifted away from the distracter, so that by 1400 ms after the word began, children initially fixating the distracter had shifted to the target about 90% of the time. The lowest curve (unfilled circles) displays children's responses on T-onset baseline trials. Children shifting away from the target raise the curve. Most of the time, children did not shift away from the target; this curve never exceeds 25%. These results for baseline trials replicate previous findings (e.g., Fernald et al., 1998; Swingley et al., 1999). The center line (filled triangles) displays responses on mismatch trials. Because there is no target or distracter on mismatch trials, all mismatch trials on which children were fixating either picture are plotted. Shifts away from this initial picture to the other picture raise this curve; shifts back lower it.

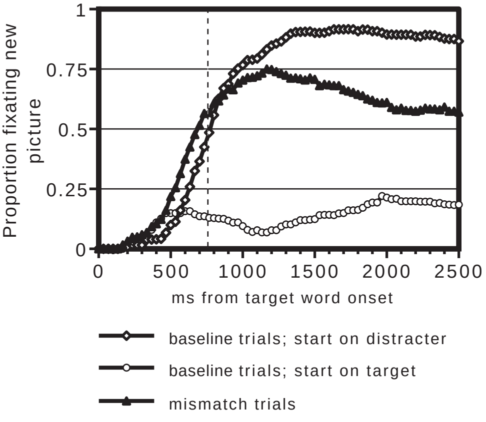

FIG. 1. Results of Experiment 1, showing children's eye movement responses over time while hearing a familiar pictured word (baseline trials) and while hearing a familiar but unpictured word (mismatch trials). The *x*-axis shows time, starting from the onset of the target word. The dotted vertical line indicates the average offset of the target words. The uppermost curve (unfilled diamonds) represents shifts from distracter to target on baseline trials. The lowest curve (unfilled circles) represents shifts from target to distracter on baseline trials. The middle curve (filled triangles) represents shifts from the initially fixated picture to the other picture on mismatch trials.

Figure 1 shows that children's initial responses on baseline D-onset trials and on mismatch trials were very similar: in general, children hearing a label that did not match the currently fixated picture swiftly shifted to the other picture—whether that other picture corresponded to the spoken word or not. The figure also shows that on mismatch trials, although children shifted rapidly, they did not then continue to fixate the new picture for long; after about 2 s following the onset of the target word, children's fixations were almost evenly divided between the two pictures. Children seemed to recognize that neither picture matched the target word.

Because there were no significant effects of stimulus order or of sex across measures, analyses will be collapsed over these variables.

Children tended to look at the named picture on baseline trials, as verified by analyses of target fixation proportion: children fixated the target 77.9% of the time on baseline trials, which is well above the 50% expected by chance ($t(23) = 13.5$, $p < .0001$; all reported *t*-tests are two-tailed unless noted otherwise).

Children's first responses on mismatch and baseline D-onset trials were not significantly different, as shown by analyses of shift proportion and response latency. Children's likelihood of shifting (shift proportion) was slightly higher on baseline D-onset trials (0.865) than on mismatch trials (0.792), but this difference was not significant ($t(23) = 1.6$, $p > .10$). The mean response latency on baseline trials was 808 ms, whereas the mean RT on mismatch trials was 748 ms, a difference that was not significant ($t(23) = 1.3$, $p > .10$). Thus, as children heard the target words (i.e., during the first 750 ms or so after target onset), children responded the same way on mismatch trials and baseline D-onset trials, rapidly shifting from the current picture to the other one. As suggested by Fig. 1, however, after this initial shift, children's behavior did depend on whether the newly fixated picture corresponded to the heard word. On baseline trials, the newly fixated picture was the target, and children tended to maintain their gaze to the named picture (only shifting back 4.1% of the time). On mismatch trials, of course, the newly fixated picture did not match the heard word, and children relatively frequently shifted back to the first distracter within the test period (43.3% of the time). This difference between conditions was significant (within-subjects t-test, $t(23) = 5.5$, $p < .001$). Children did not tend to shift away from the screens (i.e., look around the room) upon hearing mismatching words; this sort of response never exceeded 5% of trials in any of the reported experiments, and no condition differences in this response were significant or systematic.

Analyses of individual items showed the same pattern of results as the analyses by subjects. Response latencies by items did not differ in the two conditions; nor did shift proportion on mismatch trials and baseline D-onset trials (RT, $t(5) = 0.86$; shift proportion, $t(5) = 0.42$; both $p > .10$).

Examination of children's eye movements during the familiarization phase of each trial showed that children fixated both pictures on 90% of test trials before hearing the target word. Thus, on most trials children had presumably identified both pictures before the target word began. If children's eye movements, like those of adults, are taken to reflect lexical activation (e.g., Allopenna, Magnuson, & Tanenhaus, 1998), the results show that lexical activation was not significantly faster when children had just seen a picture of the spoken target (on baseline D-onset trials) than when children had not (on mismatch trials). Thus, there was no evidence that pictures primed their labels.

Discussion

In Experiment 1, 24-month-olds who were fixating a picture as they heard it named tended to continue fixating that picture. By contrast, when children heard a familiar word that did not match the fixated picture, they tended to shift quickly (about 775 ms) to the other picture. This initial response did not depend on whether the "new" picture, to which children shifted, matched the spoken word. This result implies that children's first responses are "go/no-go" responses based only on the fixated picture, without taking the other picture into consideration. Following this initial response, however, children's behavior was contingent upon whether the newly fixated picture matched the heard target word. If it did (i.e., on baseline trials), children almost always continued to look at it; if it did not (i.e., on mismatch trials), children were 10 times more likely to shift away again. Finally, seeing the target a few seconds before its label was spoken did not lead to facilitation in response latency, as shown by the similarity of the baseline and mismatch response latencies; thus, there was no detectable priming from the target picture to the spoken target word.

This pattern of results shows that 24-month-old children are capable of rapidly recognizing a familiar word and retrieving aspects of its meaning, even when the word designates an object that is not present. This suggests that the interpretation of speech about absent objects does not present a major hurdle for children at this age, at least for familiar words in simple sentence contexts.

However, there is another possible interpretation of these results, which is addressed in Experiment 2. This alternative is that children's eye movements do not in fact reflect the time course of word recognition. Rather, children's eye movements may reflect a task-dependent strategy, as follows. Suppose that children viewing a picture activated the phonological form of a label for that picture. Suppose also that children maintained their gaze if the heard target word matched this preactivated phonological form, and shifted their gaze if the heard target failed to match the preactivated phonological form. These assumptions about children's behavior could account for some of the results of Experiment 1. For example, suppose that a child presented with pictures of a duck and a car starts out looking at the duck and activates the sound form *duck*. If this matches the spoken target word (*duck*), the child maintains her gaze (correct T-onset performance); if this does not match the spoken target word (*car*), the child shifts (correct D-onset performance). According to this account, eye movements mirror the timing of a comparison between the sound forms of preactivated and heard words. Crucially, here children's eye movements are not related to children's understanding the meaning of the spoken target word. We will refer to this scenario as "phonological preactivation." The phonological-preactivation account does not provide an explanation for children's greater tendency to shift back to the initial distracter on mismatch trials than on baseline trials, but it does present an alternative explanation for the identical pattern of first responses on mismatch and baseline trials. If this account is correct, children's performance in Experiment 1 may not in fact be relevant to situations in which children hear speech referring to objects or actions that are not in view.

The alternative to the phonological-preactivation hypothesis is that children's eye movements reflect the timing of children's recovery of semantic information from the spoken target word. On this account, children hearing *duck* shift their fixation away from a picture of a car because *duck* is an inappropriate label for a car, not because *duck* sounds different from *car*. To evaluate two accounts, Experiment 2 assessed children's responses to nonce words.

EXPERIMENT 2

Interpretation of the previous experimental results hinges on the nature of the mental operations that gave rise to the observed fixation responses. It is clear that children did not simply shift their gaze whenever they heard *any* word in the appropriate sentence position; if this were so, children would have shifted even if they were fixating the target. However, children may have shifted upon hearing any word or potential word *other than* the name of the fixated picture. This would not require a search of the lexicon during the speech stimulus; it would only require a decision about whether the heard word was the one evoked by the fixated picture. One way to rule out this account is to show that children's fixation responses depend upon whether the spoken target word is a known word or not, which can only be determined by a search of the lexicon. Thus, if children respond differently to an unpictured word they know (the mismatch condition of Experiment 1) than to an unpictured word form they do not know, children's responses must depend on a search of their lexicon. In Experiment 2, then, the mismatch trials of Experiment 1 were replaced by trials on which children heard an unfamiliar novel target word (which we will call a "nonce word"). Performance on these nonce trials was compared with performance on baseline trials, as in Experiment 1.

Methods

Participants. The mean age of the 24 participants was 106.0 weeks (range 103.6 to 107.9). Thirteen were girls. All children were full-term births and were being raised in English-speaking households, as in Experiment 1. An additional 8 children were tested but were excluded from the final sample because they did not complete at least 15 of the 18 test trials (6), because the parent peeked under the curtain (1), or because nearby noise interfered with the test (1).

Stimuli. The visual stimuli were the same pictures used in Experiment 1, with the addition of four new filler pictures.

The speech stimuli used on baseline trials were the same as those used in Experiment 1, except that the *ball* and *duck* sentences were replaced with similar tokens in which the target words were slightly shorter (*ball*, 640 ms; *duck*, 528 ms). With this change, baseline and nonce targets had similar mean lengths (baseline, 692 ms; nonce, 681 ms). The nonce words were *bim* (633 ms), *daffle* (771 ms), *gizmo* (767 ms), *kreeb* (676 ms), *tame* (609 ms), and *trinket* (632 ms). These words were either invented or were judged unlikely to be known by 24-month-old children. They did not include any non-English speech sounds. Each of these unfamiliar words was presented once, in a *Where's the [target]* carrier. Sentences on filler trials were replaced; new sentences included *Look at the nice [kitty/birdy]* and *Look at the [kitty/birdy]*.

Apparatus and procedure. Procedural details were the same as in Experiment 1, except that the experiment consisted of 28 trials rather than 26. Six children were assigned to each of the four stimulus orders, approximately balanced by sex. Coding in Experiment 2 was completed by a single highly trained coder who was blind to the auditory stimulus and target side on each trial.

Results

The time-course of children's eye movements is pictured in Fig. 2. As this figure shows, children's responses on baseline trials were very similar to the responses elicited in Experiment 1. However, children's behavior on nonce-word trials was strikingly different from previous responses on mismatch trials. Rather than shifting to the other picture quickly and reliably, children responded more variably, often shifting slowly or inconsistently.

Once again there were no significant effects of stimulus order or of sex in any of the measures, so analyses are collapsed over these variables. Children tended to look at the named picture on baseline trials, as shown by analyses of target fixation proportion (%-to-target 78.9%, significantly greater than chance: $t(23) = 16.5$, $p < .0001$). The main comparisons of interest concerned the nonce trials and the baseline D-onset trials. Children were somewhat less likely to shift on nonce trials than on baseline D-onset trials (shift proportion, nonce, 0.69; baseline D-onset, 0.78; $t(23) = 1.5$, $p = .15$), and responses on nonce trials were significantly slower than responses on baseline trials (1009 vs 760 ms, $t(23) = 4.1$, $p < .0005$).

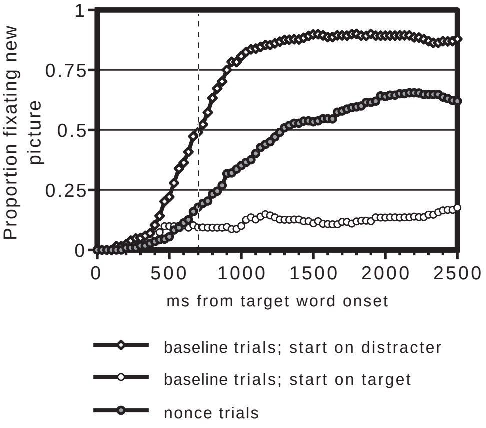

FIG. 2. Results of Experiment 2, showing children's eye movement responses over time while hearing a familiar pictured word (baseline trials) or an unknown word (nonce trials). The dotted vertical line indicates the average offset of the target words. Each curve shows the proportion of trials on which children were fixating a picture different from the picture they had been fixating at target onset, for each condition. Baseline trials are divided into two groups according to children's initial fixation position.

Analyses of individual items revealed significant effects of condition. Children were less likely to shift away from the distracter when hearing a nonce word than when hearing a familiar word (unpaired $t(10) = 3.4$, $p < .001$). Shift proportions were lower for 5 of the 6 nonce words than for any of the six familiar words. Children were slower on nonce trials than on baseline trials (unpaired $t(10) = 4.1$, $p < .005$). Children responded more slowly to all six nonce words than to any of the familiar words.

The differences between children's responses to nonce words in Experiment 2 and to familiar mismatching words in Experiment 1 were unlikely to be due to between-group differences in the ability to perform in the task, as shown by direct statistical comparisons of children's performance on baseline trials across the two experiments. In the two studies, children were equally likely to fixate the target (target fixation proportion $t(46) < 0.4$, ns), and children shifted equally quickly (response latency, $t(46) < 0.9$, ns). Figure 3 shows the response latency results for both conditions of the two experiments, il-

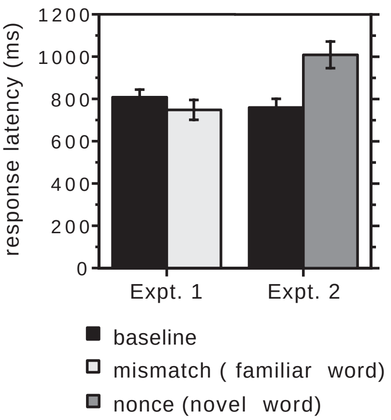

FIG. 3. Comparison of response latencies in Experiments 1 and 2, split by condition. Error bars are standard errors.

lustrating the similarity of the baseline responses across experiments and the significant delay in children's responses to the nonce words of Experiment 2.

Discussion

In Experiment 2, children hearing unfamiliar words responded slowly and irregularly, in contrast to children hearing familiar mismatching words in Experiment 1. This result is inconsistent with the assumption that children's rapid eye movements away from the initially fixated picture reflect a decision that the heard word does not match a phonological form activated by that picture. If this "phonological preactivation" account were correct, children should have shifted quickly on nonce trials, because the nonce words did not match the phonological form putatively activated by the fixated picture. Instead, we found that children hearing nonce words shifted slowly. We suggest that a rapid eye movement away from the initially fixated picture reflects a decision that the *denotation* of the heard word does not match the pictured exemplar. This result reinforces the conclusions drawn from Experiment 1: when 24-month-olds heard familiar words, activation of the meaning of those words proceeded rapidly even when the referents were not part of the visual context.

These conclusions must be tempered by a concern that is addressed in Experiment 3. Recall that the words used on mismatch trials in Experiment 1 were always words that had previously been used as targets on baseline trials. This raises the possibility that children's rapid responses were based on selection from a limited set, viz. the words previously heard in the experiment, rather than from the lexicon more generally. Thus it may be that the rapid word recognition responses seen in these experiments are only possible when either (a) the two potential referents are pictured (as in the baseline trials) or (b) the spoken target has been primed by presentation on previous trials. In order to address this issue, and to replicate the effects shown in Experiments 1 and 2, Experiment 3 was conducted with an additional group of 24-month-olds.

Experiment 3 compared mismatch, nonce, and baseline trials *within subjects*. Trial orders were modified so that half of the mismatch trials used spoken words that had not occurred on previous trials, and that had not been pictured on previous trials. The other mismatch trials used words that had served as targets on previous trials.

EXPERIMENT 3

In Experiments 1 and 2, children responded in three different ways: sustained gaze (when hearing a familiar label naming the fixated picture), rapid shifting (when hearing a familiar label naming an object other than the fixated picture), and slow shifting (when hearing a nonce word). These results indicate that children's behavior depends on their knowledge of the meaning of the spoken target word. Experiment 3 attempts to replicate these findings within subjects and assesses the potential for priming across trials.

Methods

Participants. The mean age of the 24 participants was 106.1 weeks (range 104.3 to 109.3). Half were girls. All children were full-term births and were being raised in English-speaking households. An additional six children were tested but were excluded from the final sample because they did not complete at least 15 trials.

Stimuli and procedure. Visual and auditory stimuli were taken from those used in the preceding experiments. Target words on baseline trials included *baby, ball, car, doggie, duck,* and *shoe*. The experiment consisted of 20 test trials and 6 fillers. Of the test trials, 12 were baseline trials, 4 were mismatch trials (using the target words *baby, car, doggie,* and *shoe*), and 4 were nonce trials (using the targets *bim, daffle, gizmo,* and *tame*). All targets on test trials were presented in *Where's the. . .* carriers. The number of mismatch and nonce trials was reduced to four each (as opposed to six in the preceding experiments) to keep the ratio of baseline trials to "strange" trials reasonably large, so as to not bewilder the children.

Four stimulus orders were created. The third and fourth orders used the same ordering of pictures as the first and second, but the named targets on baseline trials were reversed. Each half of the experiment included 6 baseline trials, 2 nonce trials, and 2 mismatch trials. The first-half mismatch trials used spoken words that had been neither uttered nor pictured on previous trials. The second-half mismatch trials used spoken words that had previously served as targets on baseline trials. In two orders, the "new" mismatch words (first block) were *car* and *shoe*, and the "old" mismatch words (second block) were *baby* and *dog*; in the other orders this was reversed. The counterbalancing constraints that held in the previous two experiments (regarding target side, picture order, etc.) held in the third as well. Coding was completed by a single highly trained coder who was blind to the auditory stimulus and target side on each trial.

Results

Using a within-subjects design, we found that the time-course of children's eye movements in response to baseline, mismatch, and nonce words in Experiment 3 was very similar to that found in the two previous experiments (see Fig. 4). There were no significant effects of sex or stimulus order across measures, so analyses will be collapsed over these variables.

First, confirming that children recognized words on baseline trials, target fixation proportion significantly exceeded 50% (mean proportion by subjects, 0.783; $t(23) = 12.5$, $p < .0001$).

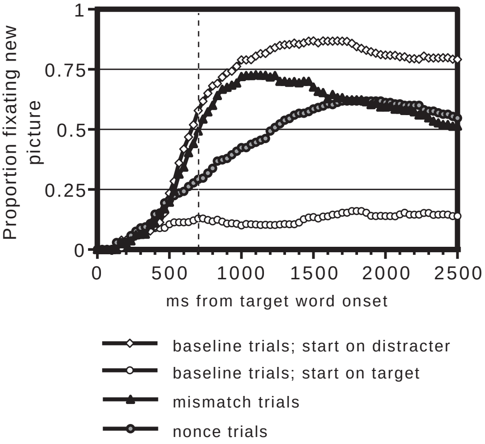

FIG. 4. Results of Experiment 3, showing children's eye movement responses over time as children heard familiar pictured words (baseline), familiar unpictured words (mismatch), and unfamiliar words (nonce). The dotted vertical line indicates the average offset of the target words. Each curve shows the proportion of trials on which children were fixating a picture different from the picture they had been fixating at target onset, for each condition. Baseline trials are divided into two groups according to children's initial fixation position.

If children's initial responses to familiar words are based upon the retrieval of these words from the lexicon, children would be expected to shift their fixation quickly and reliably on mismatch trials and D-onset baseline trials, but not on nonce trials and T-onset baseline trials. These predictions were upheld. One-way ANOVAs comparing conditions (mismatch, nonce, baseline D-onset) revealed significant effects, both for response latency ($F(2,36) = 7.2$, $p < .005$) and shift proportion ($F(2,44) = 11.1$, $p < .0001$).[4] These effects were driven by differences between the nonce trials and the other two types of trials. Children were significantly less likely to shift on nonce trials than on baseline D-onset trials (shift proportion, $t(22) = 6.3$, $p <$

---

[4] Not all children contributed response latency or shift proportion data for all three conditions, as is reflected by the varying degrees of freedom in the *F* tests. If a child failed to shift on a D-onset trial, no RT was generated; if a child happened to be fixating neither picture at target onset, no shift proportion value was generated. With only four trials in some conditions, it was inevitable that some children would be excluded from some analyses.

.0001) or mismatch trials ($t(22) = 2.8, p < .01$). Children also responded significantly more slowly on nonce trials than baseline trials (RT, $t(23) = 3.1, p < .01$) or mismatch trials ($t(18) = 2.4, p = .025$). However, children shifted equally often, and equally quickly, on mismatch and baseline trials (mean shift proportion difference, 0.032, $t(23) = 0.5, p > .20$; mean RT difference, 57 ms, $t(21) = 1.0, p > .20$). Moreover, children's rapid initial shifts on mismatch trials were often followed by shifts back to the picture they had been fixating initially, suggesting that the children were continuing to search for the referent of the target word. Of those mismatch trials on which children shifted from one picture to the other, they then shifted back 34.1% of the time; the proportion of these second shifts on baseline trials was only 9.1%, a significant difference (within-subjects *t*-test, $t(22) = 3.3, p < .005$). Thus, Experiment 3 closely replicated the effects seen in Experiments 1 and 2.

Further analyses compared mismatch and baseline responses in the first and second blocks of the experiment, to evaluate any priming effects in the mismatch trials. If children's rapid responses to mismatching words were only possible when those words had been primed by previous exposure in the experiment, responses in the mismatch condition should have been slower in the first block than in the second block. This effect was not found. Two-way ANOVAs (block X condition)[5] did not reveal any significant effects of experiment half, nor any interactions between condition and experiment half, in response latency, shifting proportion, or target fixation proportion (all $F \leq 1.3$, ns). For example, children were nonsignificantly faster to shift on mismatch trials in the first half than in the second half (first, 695 ms; second, 790 ms), a tendency in the wrong direction for the priming hypothesis; children were also nonsignificantly less likely to shift in the first half than in the second (first, 69%; second, 76%).

Discussion

Experiment 3 replicated within subjects the findings of Experiments 1 and 2. When children heard a word that matched the picture they were looking at, they tended to keep looking at that picture. When children heard a word that did *not* match the picture they were looking at, children's responses depended on whether the spoken target was a familiar word or an unfamiliar word: familiar words elicited fast and consistent eye movements, and unfamiliar words elicited an inconsistent mixture of slow shifts, fast shifts, and maintenance of gaze. Experiment 3 also extended the mismatch condition to include trials on which the targets had not been heard (or pictured) on preceding trials. Even for these trials, children responded quickly, suggesting that responses in this task draw upon children's lexicons, and not just upon the subset of words that (potentially) have been primed on previous trials. Children rapidly recognized words they had never heard spoken, nor seen pictured, in the test situation.

These results suggest that by 24 months, words rapidly give rise to ideas in children's minds. Like adults, children attending to speech quickly recognize words and activate the meanings of those words, even when the words' referents are neither available in the immediate environment nor relevant to the preceding discourse. This is shown by children's rapid responses on the mismatch trials of Experiments 1 and 2. Although children could in principle shift their gaze whenever the sentence failed to contain the phonological form suggested by the currently fixated picture, they do not appear to do so.

Our failure to detect any effects of priming in children's responses to new and recently encountered words does not indicate that children lack the mechanisms that underlie priming in adults. In fact, long-term auditory priming of familiar words has already been shown in slightly older children. In one study (Church & Fisher, 1998), 26-month-olds first heard a single instance of several different words, and then, a few minutes later, they were asked to correctly

---

[5] Because the comparison of interest was between mismatch trials in the first and second halves, compared against baseline-trial performance, the nonce condition was not included in these analyses. However, the results of the ANOVAs were the same when all three levels of Condition were included.

repeat words distorted through moderate low-pass filtering. Children's performance was significantly better for familiarized words than for new words. Adults and older children showed similar effects. Indeed, we suspect that priming effects in word recognition could be shown using the method employed in the current experiments as well, and we do not wish to overstate our null priming results here. The important conclusion to draw is that even when words could *not* have been primed, children's recognition of these words was still fast and reliable.

The fact that the children in Experiment 3 responded differently to nonce and mismatch trials, while still performing well on baseline trials, indicates that the results of the first two experiments cannot be attributed to differences between the subjects of Experiments 1 and Experiment 2. Rather, the pattern of responses to mismatch and nonce words reflects the nature of task: when a sound pattern corresponds to a word in the child's lexicon, the child evaluates whether that word's meaning is consistent with the fixated picture. If the match is adequate, children maintain their fixation; if it is inadequate, children shift quickly. But if the word is unknown, children show an inconsistent pattern of responses.

GENERAL DISCUSSION

These experiments show that children's processing of words in speech is *decontextualized* in the sense that children recognize words quickly even in the absence of the words' referents. We suggest that children attending to familiar words in speech rapidly activate the meanings of those words. This claim derives from the similarity of children's initial eye movement responses when target words were pictured (baseline D-onset trials) and when they were not (mismatch trials). When children heard a word they knew and that word did not match the picture they were fixating, children shifted rapidly; when that word did match the fixated picture, they tended to maintain their gaze.

This pattern of results contradicts the "phonological-preactivation" account of children's eye movement responses in this task. If children simply shifted their gaze upon hearing a target word different from a phonological form activated based on the picture, children would have shifted equally in the mismatch and nonce conditions. We argue that the equivalence of the mismatch and baseline responses and the observed difference between the mismatch and nonce responses indicate that children's eye movement behavior reflects the presence or absence of lexical activation. As children hear the target word, phonetically matching lexical candidates are activated. The time course of this activation is reflected in children's eye movements: when children recognize that the heard word does not refer to the fixated picture, they tend to move their eyes to the other picture rapidly and reliably. They do this because they are looking for a referent of the spoken target word. On the other hand, when children recognize that the heard word does refer to the fixated picture, they are most likely to maintain their gaze.

One might argue that a more complex version of the "phonological-preactivation" account could explain the results of the three experiments without implicating rapid semantic activation. This account assumes that children's fixation of a given picture results in activation of that picture's name. This activation persists while the child continuously compares the heard target word with the forms in the child's lexicon. As long as no match is found between the heard word and a word the child knows, the child is slow and inconsistent in rejecting the fixated picture and shifting to the alternative. If the heard word *does* match a known word, the child then compares this word with the preactivated sound form. In the case of a match, the child maintains fixation; in the case of no match, the child shifts quickly. This argument implies that children in effect perform a continuous lexical decision task, the outcome of which determines whether they will shift upon hearing a phonological mismatch with the preactivated word.

Although this account is consistent with some of our results, we find it implausible. First, it does not predict our finding that children shifting their fixation on mismatch trials frequently shifted *back* to the original picture when they discovered that their first shift did not take them to a picture matching the heard word. This result is expected only if children hearing *dog* are looking for a canine. Second, in some previous studies we have used a picture of a doll that was referred to either as *baby* or as *doll*, depending on the experiment. If children preactivated "baby" or "doll," they should shift away upon hearing "doll" or "baby"; across experiments, poor performance would be expected. However, children's performance is as good with this picture as with other pictures, suggesting that children fixate it because they consider it a reasonable referent for either word. Thus, while it is possible to tailor an elaborated "phonological preactivation" account to fit some of the existing results, in our view this account is not well motivated. On the other hand, all of the results discussed can be explained in a more straightforward way by assuming that children's responses are driven by semantic aspects of the heard word.

We acknowledge that the nature of these semantic aspects is not clear. Hearing a word like *doggie* evidently raises in children's minds some notion of [canine] that is sufficient to eliminate such things as balls and shoes as potential referents. We have discussed this in a general way as activation of the meaning of the word *doggie*, but even a very limited "meaning" would have been sufficient in the present task; we cannot claim, for example, that hearing the word *doggie* necessarily gave rise to a rich, sensuous notion of dogs. It also seems likely that the pictorial nature of the task led children to activate pictorial features of dogs, as opposed to, say, knowledge of how dogs sound or feel. This is simply a reflection of the broader principle that word meaning is to some degree constructed by the listener in interpreting an utterance, and in this sense, complete "decontextualization" is impossible (e.g., Johnson-Laird, 1987; see also Sedivy, Tanenhaus, Chambers, & Carlson, 1999, regarding the implications of this fact for online sentence comprehension). However, the role of context in leading listeners to consider only the most relevant aspects of word meaning is quite different from the role context appears to have in infancy. A 13-month-old who shows no sign of recognizing a word like "cookie" unless there is a cookie in front of him may not retrieve *any* semantic information linked to the word "cookie." This may be contrasted with the older child studied by Savage-Rumbaugh et al. (1993) leaving the room to retrieve a potato on command. Clearly this child activated some relevant notion of "potato" to guide her search. Our results suggest that this activation proceeds rapidly and efficiently as words are heard, in much the same way as semantic activation proceeds in adults, though slightly less quickly (Swingley et al., 1999).

The use (or nonuse) of visual context in word recognition may be different from the use of context in other language comprehension processes. It would be premature to generalize from the present results to, for example, children's understanding of syntactic structure. We found that children who had just seen a dog did not recognize *doggie* more readily than children who had not; but children shown a dog chasing a cat might be aided in the interpretation of *The dog chased the cat* but not *The cat chased the dog.* The mechanisms underlying the integration of contextual and syntactic information are not well understood even in adults with mature linguistic systems; study of the analogous problem in young children has barely begun (see, e.g., Trueswell, Sekerina, Hill, and Logrip, 1999, for discussion). Thus, while children can recognize words rapidly and efficiently without being cued by objects or pictures, this may or may not be true of other aspects of speech comprehension.

It also seems reasonable to suppose that the developmental time course of decontextualization differs in comprehension and production. Children capable of recognizing words referring to unseen objects might nevertheless find it relatively difficult to name those objects without visual cuing. A recent study by Dapretto and Bjork (2000) provides an instructive example. Children ranging from 14 to 24 months played a game involving four pictures of objects whose names the children could say, as assessed by parental report. With the children watching, an experimenter hid two of the pictures in a plain opaque box and the other two pictures in a similar box that had copies of all four pictures displayed on one side. After each hiding event, children were asked to name the contents of the box. Children were substantially more successful in naming the objects in the "cuing" box with the four pictures on it than in naming the objects in the plain box. The authors argue that the locus of the cuing effect is in lexical retrieval; although children knew what was in each box, they were better able to retrieve and produce words for visually cued objects than for absent objects. The size of this effect did not vary significantly with children's vocabulary size, suggesting that similar effects might be found in the slightly older children we studied here.

This dissociation between word recognition and production would provide further evidence that decontextualization is not a unitary development which children either have or have not undergone. Just as an infant might be context-bound in understanding some words and decontextualized in understanding others (Snyder et al., 1981), young children are likely to show dependence on visual supporting context in some situations or tasks and not in others. In our studies, we have attempted to reduce extraneous task demands as much as possible, while using sentences of the sort parents often use in speaking to children. Under these conditions, children do not seem to rely on visual context in the rapid understanding of familiar object labels.

The present findings also help to clarify previous results obtained using visual fixation procedures. As described in the Introduction, studies by Fernald, Swingley, and Pinto (in press) and Swingley, Fernald, and Pinto (1999) suggested that young children's interpretation of speech is *continuous* in the sense that incoming speech incrementally modulates the activation of words in the lexicon. In Fernald et al. (in press), for example, children heard whole words (e.g., *baby*) and partial words (*bay*—). As children heard the words, they rapidly shifted their gaze away from the distracter picture even when only the first part of the word was presented. In a related study, Swingley et al. (1999) found that children hearing a word like *doggie* continued to look at a distracter picture of a doll about 300 ms longer than they looked at a distracter picture of a tree. The delay in rejecting the doll picture was comparable to the duration of the phonetic overlap in the words *doll* and *doggie*. The "phonological-preactivation" explanation for both of these findings is that children viewing a distracter picture in each experiment activated the phonological form of a name for that picture and shifted their gaze when the spoken target word became inconsistent with this preactivated form.

The present results render this account unlikely; children the same age, tested with an identical procedure, do not show evidence of this "phonological preactivation" strategy. Rather, we suggest that children hearing the target words in these experiments activate the words in their lexicons that correspond phonetically to the speech. This activation of the sound-forms of words in the lexicon gives rise to activation of semantic knowledge, which is the basis for children's eye movement behavior. In the Swingley et al. (1999) study, children hearing the first 300 ms of *doggie* initially activated both "doggie" and "doll," yielding continued fixation to the doll, but not to the tree. When *do*. . . became *dog*. . ., however, "doll" was no longer activated, leaving only "dog," and directing children away from the picture of the doll. Thus, children's delay in shifting away from the doll relative to the tree is readily explained by the same type of multiple-activation account that is used to explain similar effects in adults (e.g., Marslen-Wilson, 1987). Such effects were also found using the same stimuli in a control experiment by Swingley et al. (1999) and in an analogous task with more complex visual arrays (e.g., Allopenna et al., 1998).

We suggest that there is substantial continuity in the basic processes that underlie word recognition in adults and in children at least as young as 24 months. Like adults, children process speech rapidly and continuously, interpreting the meanings of words as the words unfold over time, and as in adults, the activation of words is reduced when words are mispronounced (Swingley & Aslin, 2000). The present results reveal yet another way in which speech processing by young language learners is similar to that of adults. Consistent with Tiedemann's observations two centuries ago, we found that by 24 months, familiar nouns rapidly evoke ideas not exemplified in the child's immediate environment.

REFERENCES

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development,* **67,** 635–645.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language,* **38,** 419–439.

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

Baldwin, D. A. (1993a). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology,* **29,** 832–843.

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

Bates, E., Camaioni, L., & Volterra, V. (1979). The acquisition of performatives prior to speech. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 111–129). New York: Academic Press.

Bloom, L. (1973). *One word at a time: The use of single-word utterances before syntax*. The Hague: Mouton.

Bloom, L. (1993). *The transition from infancy to language*. Cambridge: Cambridge Univ. Press.

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life. *Monographs of the Society for Research in Child Development,* **62** (2), serial no. 250.

Church, B. A., & Fisher, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language,* **39,** 523–542.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America,* **64,** 44–56.

Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *Journal of Verbal Learning and Verbal Behavior,* **19,** 297–315.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language,* **32,** 193–210.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken word recognition in French. *Journal of Memory and Language,* **42,** 465–480.

Dapretto, M., & Bjork, E. L. (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth. *Child Development,* **71,** 635–648.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development,* **59** (5), serial no. 242.

Fernald, A., McRoberts, G. W., & Swingley, D. (2001). Infants' developing competence in recognizing and understanding words in fluent speech. In J. Weissenborn & B. Hoehle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic, and neurophysiological aspects of early language acquisition* (Vol. I, pp. 97–123). Amsterdam: Benjamins.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the second year. *Psychological Science,* **9,** 228–231.

Fernald, A., Swingley, D., & Pinto, J. P. (in press). When half a word is enough: Infants can recognize spoken words using partial acoustic–phonetic information. *Child Development.*

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language,* **14,** 23–45.

Haith, M. M., Wentworth, N., & Canfield, R. L. (1993). The formation of expectations in early infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 8, pp. 251–297). Norwood, NJ: Ablex.

Huttenlocher, J. (1974). The origins of language comprehension. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 331–368). Potomac, MD: Erlbaum.

Johnson-Laird, P. N. (1987). The mental representation of the meaning of words. *Cognition,* **25,** 189–211.

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science,* **277,** 1984–1986.

Lewis, M. M. (1937). The beginning of reference to past and future in a child's speech. *British Journal of Educational Psychology,* **7,** 39–56.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition,* **8,** 1–71.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance,* **15,** 576–585.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken-word recognition. *Cognition,* **25,** 71–102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology,* **18,** 1–86.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of dependence between retrieval operations. *Journal of Experimental Psychology,* **90,** 227–234.

Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior,* **2,** 217–228.

Naigles, L. G., & Gelman, S. A. (1995). Overextensions in comprehension and production revisited: Preferential looking in a study of dog, cat, and cow. *Journal of Child Language*, **22,** 19–46.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America*, **87,** 2705–2715.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52,** 189–234.

Sachs, J., & Truswell, L. (1978). Comprehension of two-word instructions by children in the one-word stage. *Journal of Child Language*, **5,** 17–24.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, **58** (3–4), serial no. 233.

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

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, **71,** 109–147.

Snyder, L. S., Bates, E., & Bretherton, I. (1981). Content and context in early lexical development. *Journal of Child Language*, **8,** 565–582.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, **76,** 147–166.

Swingley, D., Pinto, J. P., & Fernald, A. (1998). Assessing the speed and accuracy of word recognition in infants. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 257–277). Stamford, CT: Ablex.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, **71,** 73–108.

Taine, H. (1877). Acquisition of language by children. *Mind*, **2,** 252–259.

Tiedemann, D. (1927). Tiedemann's observations on the development of the mental faculties of children (C. Murchison & S. Langer, Trans.). *Pedagogical Seminary and Journal of Genetic Psychology*, **34,** 205–230. [Original work published 1787]

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

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

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, **73,** 89–134.

Walley, A. C. (1987). Young children's detections of word-initial and -final mispronunciations in constrained and unconstrained contexts. *Cognitive Development*, **2,** 145–167.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, **32**, 25–64.

(Received June 30, 2000)
(Revision received February 9, 2001)