Combining Techniques to Reveal Emergent Effects in Infants' Segmentation, Word Learning, and Grammar

George Hollich

Purdue University

Key words

domain-general

grammar

methodology

speech perception

word learning

Abstract

This paper provides three representative examples that highlight the ways in which procedures can be combined to study interactions across traditional domains of study: segmentation, word learning, and grammar. The first section uses visual familiarization prior to the Headturn Preference Procedure to demonstrate that synchronized visual information aids in speech segmentation in noise. The second section uses audio familiarization prior to the Preferential Looking Procedure to demonstrate that speech perception aids in the learning of meaning. The third section uses visual familiarization prior to the Preferential Looking Procedure to demonstrate that attentional distractions inhibit grammatical understanding. Thus, what infants see affects what they hear. What infants hear affects the words they learn. What infants remember affects the sentences they understand.

1 Introduction

Acquiring language is a multifaceted task. Infants must segment the words in the speech stream, map meaning to those words, and discover the rules for combining words to express new meanings. Furthermore, these domains of segmentation, word learning, and grammar are likely not independent from each other nor other aspects of children's development, including the growth of social skills, attention, and working memory. For language, versions of this interactionist perspective have been tested and supported by many (Baldwin & Tomasello, 1998; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Smith, 1999; Waxman & Booth, 2001; Woodward & Markman, 1998). This perspective is also consistent with more general models of cognitive development (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Ford & Lerner,

1992; Thelen, & Smith, 1994). This paper adds to this growing literature by providing three representative examples that highlight the ways in which procedures can be combined to study interactions across traditional domains of language acquisition study: segmentation, word learning, and grammar.

The studies reported combine the Headturn Preference Procedure and the Preferential Looking Procedure. In the Headturn Preference Procedure, infants are familiarized to an auditory pattern and then tested on their memory for that pattern. In the Preferential Looking Procedure, infants see two objects presented on a screen (or on 2 monitors) while the audio requests that they look at one of the objects. Combining these procedures produces techniques flexible enough to familiarize infants with acoustic information to test for auditory effects on word learning or to familiarize infants with visual information to test for visual effects on speech perception and grammatical understanding. For example, speech perception is generally considered in purely acoustic terms, but the first section demonstrates that vision matters; familiarizing seven- to eight-month-olds with synchronized visual information prior to the Headturn Preference Procedure aided in speech segmentation. Likewise, word learning usually focuses on the word learning situation itself, but the second section demonstrates that prior exposure to the acoustic form matters; familiarizing 24-month-olds with new words prior to the Preferential Looking Procedure aided the learning of their meaning. Finally, the acquisition of grammar is often considered in modular terms, separate from the rest of development. The third, and final, section demonstrates that children's performance on a syntactic understanding task is profoundly affected by visual attention; familiarizing 20-month-olds with complex visual scenes inhibited the grammatical understanding they had previously demonstrated in the Preferential Looking Procedure.

2 Segmentation

One of the first tasks faced by infants is to segment individual words from the speech stream. Speech is not punctuated with spaces, commas, or periods. Less than 7% of the speech directed at infants consists of isolated words (van de Weijer, 1998); furthermore, this low percentage occurs even when mothers are explicitly directed to teach their children individual words (Aslin, Woodward, LaMendola, & Bever, 1996; Newman, 2003). Nonetheless, research conducted over the past ten years has demonstrated that by the time they are eight months of age, infants are capable of extracting meaningful units from the acoustic stream. Using the Headturn Preference Procedure, Jusczyk and Aslin (1995) familiarized 7.5-month-old infants to fluent speech passages containing two target words (either cup and dog, or bike and feet). Infants were later tested on their preference for those words when played in isolation versus the two words with which they had not been familiarized. Infants listened longer to words contained in the stories than to words that had not occurred in those stories. These results suggest that infants had segmented and remembered the target words. The Headturn Preference Procedure has not only been used to demonstrate that infants can segment the speech stream but also to establish some of the acoustic cues that infants use in doing so. These cues include statistical cues (Saffran, Aslin, & Newport, 1996), metrical stress cues (Jusczyk, Cutler, & Redanz,

1993; Morgan, 1996), and phonotactic cues (Jusczyk, Luce, & Charles-Luce, 1994). Thus, the Headturn Preference Procedure has proven a reliable and valuable method for examining infants' abilities to segment the speech stream.

Complicating the segmentation problem, infants must also be able to follow one speech stream over another. Consider an infant sitting in a room with her family; her mother might be speaking while her older sister is watching television and her two brothers are arguing nearby. In order to understand her mother, the infant must be able to separate her mother's speech from that of the other voices in her environment. Properly segregating voices is thus a critical first step to segmentation. Infants cannot segment what they cannot separate from distractions. Newman and Jusczyk (1996) explored infants' segmentation while segregating voices using the same stimuli as Jusczyk and Aslin (1995). However, in the Newman and Jusczyk (1996) study, the familiarization passages were blended with speech from a distractor voice. They found that infants were able to segment words only when these familiarization passages were 10 decibels (dB) more intense than the simultaneous distractor voice. Similarly, Nozza, Rossman, Bond, and Miller (1990) demonstrated that infants' ability to recognize a /ba/ /ga/ distinction in white noise was successful only above an 8 dB signal-to-noise ratio. Together, these studies suggest infants find it nearly impossible to process fluent speech at signal-to-noise ratios less than 8 dB, when speech-specific auditory cues are more difficult to perceive.

To date, research on speech segmentation has focused on the cues within the speech stream itself; however, even four-week-olds can direct their attention based on synchronization between sight and sound (Bahrick, 2001; Bahrick, Lickliter, & Flom, 2004). Dodd (1979) found that 10- to 16-week-olds will direct their attention to speech presented in synchrony with a dynamic video of that speaker's face compared to a video that is not synchronized (see also Lewkowicz, 1994; Pickens, Field, Nawrocki, Martinez, Soutullo, & Gonzalez, 1994). Infants can also use what they see on the face to recognize vowels (Kuhl & Meltzoff, 1982) and distinguish consonants (Massaro, 1998). Such skill at combining visual information with auditory information would seem especially useful for separating speech streams. That is, while speech segregation is usually considered an acoustic task, infants may use the synchronized information between a talker's face and their voice to help them focus on a target speech stream. Infants might even be able to segment that stream at levels of distraction previously impenetrable.

Recently, we (Hollich, Newman, & Jusczyk, 2005) tested whether infants would use auditory visual synchronization to aid speech segregation and segmentation in a series of four studies. The studies combined visual and auditory familiarization with the standard Headturn Preference Procedure. In all of the studies, 7.5-month-old infants were familiarized with a video display accompanied by an audio track of a blended stimulus. This blended stimulus consisted of the female talker (reading the passages from Jusczyk & Aslin, 1995) and a male talker (reading a distractor passage). Importantly, the target audio (the female speaker) was the same average loudness as the distractor audio (the male speaker). This target stimulus was thus 10 dB softer than the level at which infants had been shown to successfully segment a speech stream using audio cues alone. We expected that the addition of video information — specifically synchronized visual information — could help infants succeed in this task. In all four studies, after visual familiarization, we then tested

infants' memory for target words using the standard Headturn Preference Procedure. In this manner, we combined visual familiarization with the standard procedure to assess visual effects on segmentation.

Each study differed in the type of visual familiarization. The first study showed a synchronized video of a female speaker. It was expected that if the synchronized video was useful to direct and focus infants' attention to the correct speech stream, then infants should succeed at levels of noise previously impenetrable. The second study familiarized infants with an oscilloscope pattern (a wiggling yellow line on green background) a that was also synchronized to the target speech stream. The logic was to test whether amodal synchronization of any kind would be enough to help infants succeed in the task. The third study was a control that familiarized infants with a static picture of the female face. We expected that the lack of visual synchronization would cause infants to fail in this version of the task, although it was possible that attentional interest in faces might have heightened infants' persistence in the task and allowed them to succeed. The final study, also a control, familiarized infants with an unsynchronized video display of the female talker speaking the passages. Again, infants were expected to be unable to segment the speech in this condition, but this condition insured that any effects seen could not be the result of increased attention due to moving video.

2.1
Method

2.1.1
Participants

Participants were 120 infants (30 in each visual condition) with a mean age of seven months 15 days (range: 7m 2d – 7m 28d); 10 additional participants were excluded as a result of fussiness. All participants were recruited using mass mailings and were from monolingual, English-speaking homes.

2.1.2
Stimuli

To create the familiarization stimuli, a video recording was made displaying a close-up of the face of a Caucasian female speaker of American English reading four passages in infant-directed speech.[1] As in Newman and Jusczyk (1996), and Jusczyk and Aslin (1995), each fluent passage was constructed around a target word (either cup, dog, bike, or feet). Distractor passages consisted of a male speaker reading the Methods section of the original Newman and Jusczyk (1996) paper. This speaker (also a native speaker of American English) was instructed to read the paper in a monotone manner, and, crucially, in adult-directed speech. The average intensity levels of the audio recordings of the male and female passages were then adjusted

---

[1]   Prior studies (Fernald & Kuhl, 1987) have shown that infants prefer to listen to the fundamental frequency modifications of infant-directed speech over that of adult-directed speech. By ensuring that the distractor voice used an adult-directed speaking style, we decreased the likelihood that infants would select that voice as the one to which they chose to attend.

using a waveform program on the computer until they were of equal RMS amplitude. The audio files were then digitally combined in such a manner that the male voice was speaking whenever the female spoke a target word. In addition, the male passage was trimmed so that the onset and offset of the passage was simultaneous with the onset and offset of the female passage.

The video of the female was then synchronized with this blended audio product to produce the video used in the first study. For the second study, the waveform of the female passages across an oscilloscope with a 30 ms running window was video-recorded and subsequently synchronized with the blended audio. This resulted in a video in which the oscilloscope was synchronized only with the female voice. For the third study, a single frame of the video was used for the duration of the familiarization. This static video provided visual information but no movement. For the final study, the video was switched for the two familiarization words, such that each child saw the video for one target while listening to the audio for the other and vice versa. In this manner, infants in this condition saw and heard exactly the same videos as in the synchronized condition, but the videos did not match the audio.

2.1.3
Procedure

The procedure consisted of a familiarization phase and a test phase, each administered in a different room and using different set-ups: one for displaying the video and the other for administering the standard Headturn Preference Procedure (Jusczyk & Aslin, 1995). During the familiarization phase, infants were seated (on their caregiver's lap) approximately 45 inches from a large presentation display. The audio was played using the built-in speakers on this display. After the infant was seated comfortably, and the parent blindfolded, the video was played through to completion regardless of infant looking. It should be noted, however, that looking times during this phase were uniformly at ceiling across all studies. All infants found the visual display to be extremely compelling. When played, each familiarization passage was 22 s long and 72 dB in average amplitude. During familiarization, a child would hear two of the passages (cup/dog or bike/feet) repeated twice (with a 1 s black screen in between) for a total familiarization phase of 88 s (44 s of exposure to each word). After the video was finished, infants and their parents were escorted into the room where the standard Headturn Preference Procedure was conducted.

In the Headturn Preference Procedure, the infant sat on the caregiver's lap in the center of a three-sided enclosure (see Fig. 1). A trial began with the flashing of the green light on the center panel. When the infant fixated on the green light, it was extinguished, and a red light on one of the side panels began to flash. When the infants' head turned at least 70 degrees towards the flashing light, the experimenter initiated the speech sample from a loudspeaker under that light. If the infant turned away from the light for a period of at least two seconds, the computer would end the trial, and the green center light would begin to flash, signaling the beginning of a new trial. Information about the direction and duration of the headturns and the total trial duration were stored in a data file on the computer. Any time the infant spent looking away was not included when measuring the total listening time. Both the experimenter and caregiver wore sound-insulated headphones that played continuous

Figure 1

The Headturn Preference Procedure. Infants sit on a caregiver's lap inside a three-sided enclosure with a flashing light on each wall. Infants' listening times test their ability to segment speech
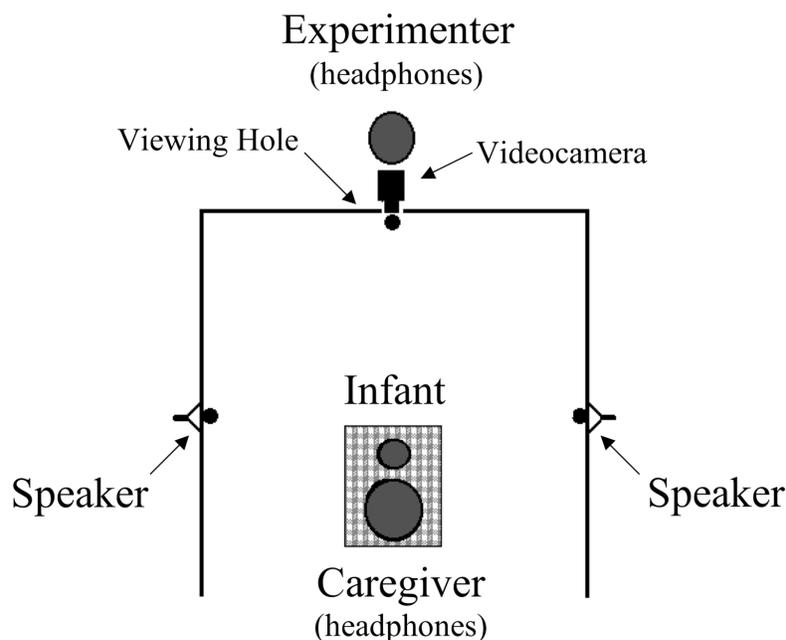
masking music to prevent them from hearing the stimulus materials throughout the duration of the experiment.

2.2
Results and Discussion

The mean looking times to the target and nontarget words are presented in Figure 2. Infants looked reliably longer at the target versus nontarget words only when they saw a synchronized display. They did so whether this display was a synchronized face, $t(29) = 4.39$, $p < .0001$, or a synchronized oscilloscope, $t(29) = 2.28$, $p < .05$. With a static display, $t(29) = 1.16$, $p = n.s.$, or with asynchronous visual information, $t(29) = 1.38$, $p = n.s.$, infants did not show evidence of looking longer at the target versus nontarget words. Thus, across four studies, 7.5-month-old infants were shown to use synchronized auditory-visual correspondences to allow them to separate two different streams of speech at signal-to-noise ratios lower than previously demonstrated. Infants were able to use a synchronized visual display to attend to and segment the fluent speech stream of a female speaker even when it was the same average loudness as that of a male distractor voice and even when the visual familiarization was one with which they were unlikely to have had any prior experience (a moving oscilloscope pattern). Infants did not succeed in this task if familiarized with a static or asynchronous video display of that speaker's face, implying that it was specifically the synchronized video

that produced the effect. In this manner, attention to synchronized visual information helps in the task of separating speech and aids segmentation.

Figure 2

Mean looking time to the target and nontarget words depending on familiarization
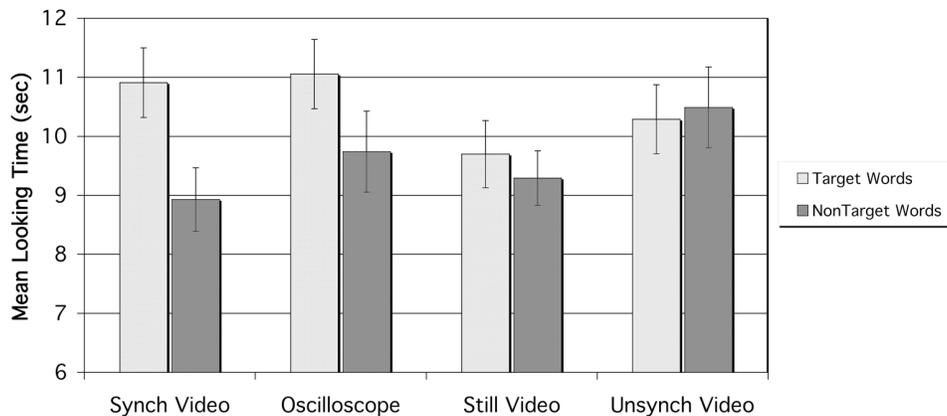
Infants may not use this visual information directly to segment speech. Even older children are poor at "speech reading" (Massaro, 1998) and the fact that infants succeeded with the synchronized oscilloscope pattern would argue against a speech-specific interpretation. Rather, it is likely that synchronized visual information directed infants' attention to the appropriate stream (see Hollich et al., 2005, for a discussion). From there, the standard cues and mechanisms came into play. Ongoing studies are examining these issues. However, regardless of the mechanism that ultimately turns out to be responsible, the current study demonstrates that visual perception can help infants to solve the speech perception task under conditions of noise. Thus, the domains of visual perception and speech perception are not as separate as they might appear. Infants can use what they see to hear better—to separate and segment the stream of speech. The next section demonstrates how the domains of speech perception and word learning interact.

3 Segmentation and word learning

Besides successfully segmenting the speech stream, infants must also learn to map the correct meaning to these newly segmented acoustic packages. As with segmentation, infants are quite skilled at this task; using the Preferential Looking Procedure, researchers have found that 12- to 17-month-olds can associate a new word with an unknown object in as few as six pairings (Hollich et al., 2000; Schafer & Plunkett, 1998; Woodward & Markman, 1998). By 24 months, infants can even learn some aspects of meaning after one repetition (Carey, 1978; Dollaghan, 1985). Despite such skills, it is likely that the domain of word learning is not independent from other domains. For example, infants can use social cues, such as pointing (Baldwin & Tomasello, 1998), or grammatical cues, such as part of speech (Waxman & Booth,

2001), to determine the meaning of a new word. Furthermore, while prior work has tended to focus principally on the word learning situation itself, it is possible that prior acoustic exposure (and infants' developing speech perception abilities) may make learning meaning easier.

To learn a word, infants must represent both the acoustic form and then make the connection to an external, possibly unfamiliar, object. Prior experience with that word spoken by that speaker might make it easier for infants to learn the word. Some suggestive evidence that hearing a particular speaker say a word may aid in subsequent learning of meaning can be found in the speech perception literature. Houston and Jusczyk (2000) have found that when infants were familiarized with one voice and tested on another, they do not show evidence of segmentation. Apparently something as innocuous as changing the speaker can impair infants' abilities to recognize a familiar word in a new context. Thus, it is possible that, even for relatively experienced word learners, skill at word learning may depend on specific experience with different talkers and on having previously heard those talkers say the target word. This section reverses the procedure from the previous section to test for such an effect. In the previous study, video familiarization was followed by the Headturn Preference Procedure. This allowed the testing of visual effects on segmentation. In the experiment reported in this section, infants were familiarized in a simulated Headturn Preference Procedure and then tested in the Preferential Looking Procedure to see if the acoustic familiarization aided word learning.

Figure 3

The Splitscreen Preferential Looking Procedure. Infants sit on blindfolded parent's lap and are taught and tested on the meanings of new words
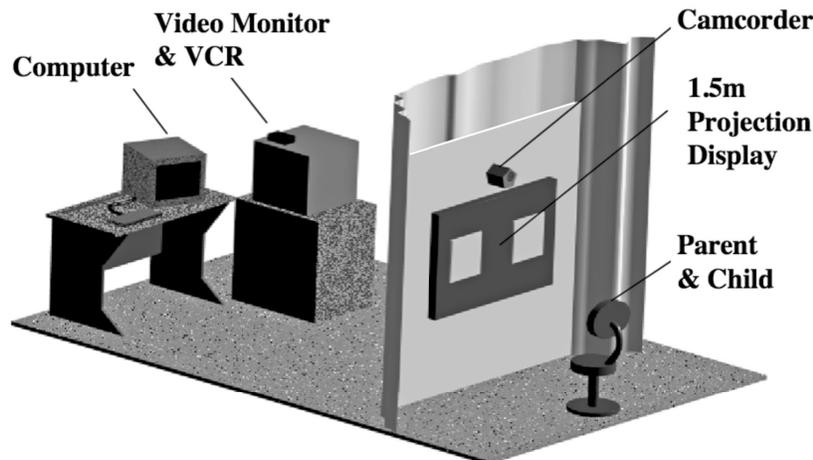
Infants first heard passages containing two unfamiliar target words — one word spoken by one talker, the other word spoken by a different talker. Next, using the Splitscreen Preferential Looking Procedure (see Fig. 3), infants were taught and tested on two new word-object pairings using these novel words. Importantly, the

training voice was one of the voices the infants had heard previously. The hypothesis was that if prior history mattered, hearing a target word spoken by a talker before learning meaning should facilitate that learning.

3.1
Method

3.1.1
Participants

Participants were 24 infants with a mean age of 23 months 15 days (range: 23 m 1 d – 23 m 29 d); two additional participants were excluded as a result of fussiness. All participants were recruited using mass mailings and were from monolingual, English-speaking homes.

3.1.2
Procedure

Table 1

Design. Auditory familiarization is followed by preferential looking to see how prior acoustic exposure affects word learning

Visual Display	Sound Track	Talker
Familiarization		
	"The nef was round and shiny…"	Talker A
	"The doop fell off the table…"	Talker B
	"The nef was round and shiny…"	Talker A
	"The doop fell off the table…"	Talker B
Training		
	"Doop, Doop, Doop"	Talker A
	"Nef, Nef, Nef"	Talker A
Test		
	"Nef, Nef, Nef"	Talker A
	"Doop, Doop, Doop"	Talker A
	"Doop, Doop, Doop"	Talker A
	"Nef, Nef, Nef"	Talker A

This study was conducted in a single room, using a simulated Headturn Preference Procedure for familiarization and the standard Preferential Looking Procedure (see Table 1) at test. For the familiarization, infants were familiarized with two sets of sentences: one set constructed around the novel word doop (/duːp/), and the other around the novel word nef (/nɛf/). Importantly, each set was spoken by a different talker. To maintain attention during this audio, a red light would flash on the screen, similar to the Headturn Preference Procedure; however, in this version the audio played through to completion regardless of infant looking.

For the preferential looking portion, this experiment involved the used of the Splitscreen Preferential Looking Procedure. This method placed children in their caregivers' laps approximately 45 inches from a large 56-inch video display. To teach infants a new word, a novel object was displayed on the screen while the audio repeated a novel label three times (e.g., "nef, nef, nef").[2] Next, a second novel object appeared and was labeled (e.g., "doop, doop, doop"). These two trials were repeated to bring the total number of labels to six for each word. To test comprehension, the two novel objects were then presented on each side of the screen. The infants then heard the label for one of the objects (e.g., "nef, nef, nef"). If the infants spent more time looking at the target (the nef) than the distracter (the doop), when the nef was requested (and not when doop was requested), this is taken as evidence that they understand the meaning of the word. All labels were counterbalanced across conditions. Caregivers wore a blindfold to ensure that they were not able to influence infants' looking behaviors.

3.1.3
Coding

All coding was conducted off-line. Videos of the participants were transferred to computer. Coders, blind to the condition being run, stepped through the videos, frame-by-frame, marking the beginnings and ends of each left and right look. These marks were then exported to a spreadsheet for analysis of the mean looking times to the target and nontarget objects. Because of the frame-by-frame nature of this process, this method was extremely accurate (to within one 30th of a second), and inter-rater reliability (as tested by random recodings of 20% of the data) was above 98%.

Figure 4

Mean looking time to the target and nontarget objects on test trials when the talker was the same or different between familiarization and training

---
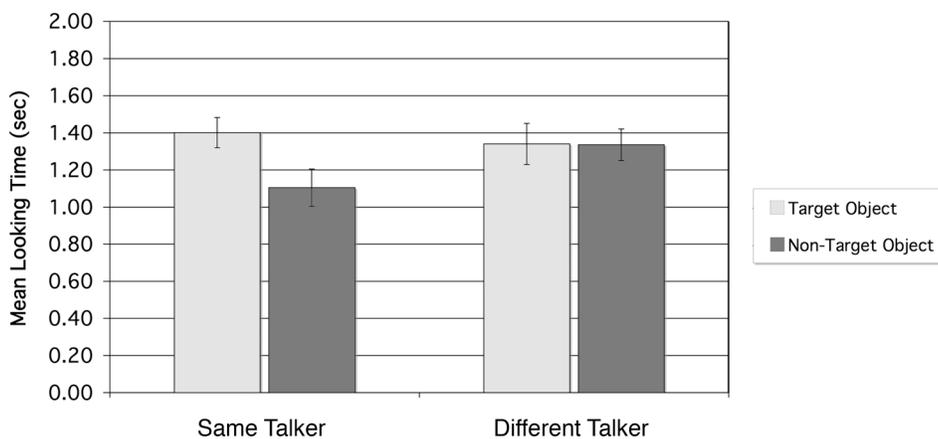
[2]   Words were repeated in isolation in order to keep training as simple as possible and to control for segmentation effects during training.

3.2
Results and Discussion

Mean looking times to the target and nontarget are presented in Figure 4. They indicate that 24-month-old infants looked longer at the target only in the case where infants had heard that talker using that word previously, $t(23) = 2.36$, $p = .02$, but not in the case where they had heard a different talker using that word during familiarization, $t(23) = 0.04$, $p = n.s.$ This effect was not the result of familiarity with the voice; infants had heard that speaker during the familiarization. This effect was not the result of familiarity with the word. Infants had heard that word during the familiarization; they just hadn't heard that speaker say that word before. In addition, this effect was not the result of different training stimuli; infants in all conditions experienced the same training. The effects observed came purely from prior exposure. To be clear, infants can learn words they haven't heard before. Follow-up studies indicated that with repetition and variation infants succeed in this task. Nonetheless, this work demonstrates that 24-month-old word learning is context-bound in a way that belies infants' apparent expertise.

Segmentation thus plays a critical role in word learning and acquisition — even in relatively mature infants. While ongoing studies are examining the specific nature of this effect, this experiment demonstrates the interactive nature of the process. As the previous section demonstrated that what infants see affects their segmentation, this section demonstrated that infants' segmentation affects their word learning — prior history matters. The final section demonstrates how attentional factors can inhibit infants' display of grammatical understanding.

4 Attentional factors and grammatical understanding

Traditional explanations propose that grammar can only be learned through the use of innate language acquisition devices (Chomsky, 1957), yet even these explanations would acknowledge that grammatical understanding (competence) and real world performance are very different. It is likely that some of the differences between these are due directly to attentional and memory issues rather than grammatical understanding per se. To decide what is innate versus what is learned, one must be able to separate out difficulties due to attentional and memory factors from those due to immature grammatical understanding. Like segmentation and word learning, a complete explanation for grammar learning will encompass multiple avenues, some specific to language and others more general.

Consider the question "Where's the flower?" Do the youngest infants truly understand the complexities of this grammatical construction, or are they initially perceiving something similar to "Blah, blah, blah … flower?" In either case, they would correctly look at the targeted object during a preferential looking task. In support of the latter, "dumb," explanation, Fernald, Mc Roberts, and Herrera (1992) found that infants only perform successfully if the target is at the end of the utterance. Likewise, the amount of familiarization appears to make a difference (Schafer & Plunkett, 1998). In addition, even five-year-olds make errors in constructing questions (e.g., "Who do

you think who's under there?" or "What do you think what Minnie tried to put on?" Thornton, 1995). It is thus possible that early on infants do not understand certain aspects of English syntactic structure — parsing only individual meanings of the words. Alternatively, it is also possible that infants' possess latent grammatical abilities that are limited by attentional and memory demands. Many grammatical tasks require that infants point to the correct picture or "act-out" the grammatical construction under study (see Hollich, Hirsh-Pasek, & Golinkoff, 1998). If this is true, then under some very simplified circumstances, perhaps using preferential looking, infants may indicate that they do understand complex grammatical constructions.

In previous work (Seidl, Hollich, & Jusczyk, 2003), we tested this hypothesis by familiarizing 15- and 20-month-olds with a vignette of an apple hitting a flower and then asking not only a baseline question, "Where is the flower?" (standard in every preferential looking task), but also "What hit the flower?" (a question that asks for the subject), and "What did the flower hit" (a question that asks for the object). Notice that the answer to the subject and object questions is the opposite of the object overtly mentioned in the question. If infants were only parsing the word 'flower' in this task, they would look at the incorrect object. We found that besides successful performance on the where questions, 15-month-olds looked significantly longer at the target in the subject questions, and 20-month-olds succeeded on both the subject and object questions. Thus, by using preferential looking with a very simple scene, which used familiar labels and objects, these results indicated a sophisticated grammatical sensitivity previously unsuspected.

Ironically, these results also created another question. If infants as young as 15 months can understand these grammatical constructions, why don't they indicate more of this understanding in their everyday performance? The answer is that while a sensitive task may reveal underlying competencies, in the real world, these competencies will remain dormant until memory and attentional abilities catch up to them. To illustrate, we recently (Hankin, Hollich, Jusczyk, & Seidl, 2002) tested how 20-month-old's performance in this task was affected when we made the task more complex, visually. Specifically, we added a third distractor object to the familiarization displays. We expected that even 20-month-olds might be distracted by the third object and as a result fail in this task, or in partial failure, might only look at the item explicitly mentioned in the question. Such behavior would be interesting because it would show that infants could understand the words and the interrogative nature of the questions but were confused about the grammatical interpretation.

4.1
Method

4.1.1
Participants

Participants were 20 infants with a mean age of 20 months 12 days (range: 19m 28d – 20m 18d); three additional participants were excluded as a result of fussiness. All participants were recruited using mass mailings and were from monolingual, English-speaking homes.

Figure 5

Visual familiarization. Infants see a flower hit an apple while a distracting object bounces on the side. Would the distraction reverse infant performance?
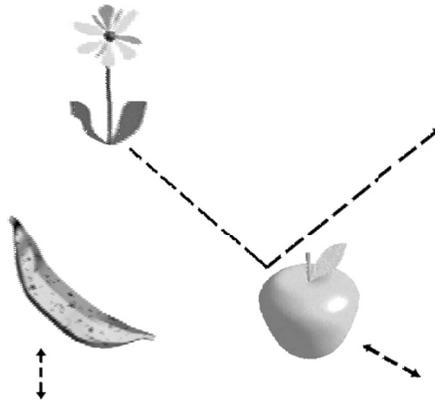
4.1.2
Procedure

Infants in this study saw short video actions depicting hitting (e.g., an apple hitting a flower) while in the corner of this video a third object (e.g., a banana) slowly bounced by itself (see Fig. 5). Infants then saw the objects involved in the hitting action side by side on the screen while the audio requested one of three possible questions: "What hit the X?" where X was the object that had been hit, "What did the Y hit?" where Y was the object that did the hitting, and "Where is the Z?" where Z was one of the objects. For example, a child might see an apple hitting a flower while a banana was bouncing on one side. Then the child would see the apple and flower together and hear "What did the apple hit?" (the object question). Next they might see a book hit a pair of keys while a ball was bouncing on one side. Then, the child would see the book and keys together and hear "What hit the keys?" (the subject question). Finally the child might see the apple and flower together again and hear "Where is the flower?" Each of these trials was repeated twice for greater accuracy, the order of trials and objects were counterbalanced across children. The apparatus and coding were exactly the same as in the study described in the last section.

4.2
Results and Discussion

Results of this study and the original study are presented in Figure 6. Recall that the original study was the same as this one except that there was no distracting third object. In the original study (Seidl et al., 2003), the 20-month-olds looked significantly longer to the target when presented with the choice between the target and the nontarget for all types of questions. However, in the present study, the 20-month-olds completely reversed their looking preference for the difficult questions; they looked only at the objects overtly mentioned in the subject questions, $t(19) = 4.82$, $p < .001$, and the object

Figure 6

Mean looking time to target and nontarget object by question type when a distracting object is added during familiarization. Original results (Seidl et al., 2003) are presented for comparison. Subject = "What hit the X?" Object = "What did the X hit?" Where = "Where is the X?"
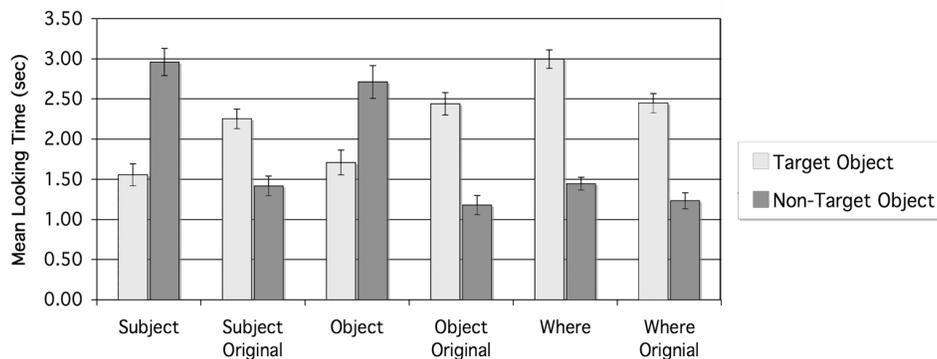
questions, $t(19) = 2.93$, $p = .009$. Thus, if asked "What hit the apple?" infants looked longer at the apple rather than the correct answer.

In this manner, addition of the third distractor object profoundly disturbed infants' ability to indicate their grammatical understanding. It appears that while grammatical understanding can be seen at 20 months, attentional factors may block infants' demonstration of this fact in practice. It is unlikely that the addition of the third object impaired grammatical understanding directly. Instead, in the current study, it is likely that something related to the visual parsing and the processing of the visual scene caused the infants' performance to reverse itself. It is possible that with three objects, infants simply cannot remember which object hit which. Perhaps the addition of the third object made the scene more difficult to understand, or perhaps infants at this age are incapable of parsing the scene correctly in the time allotted. Whichever explanation ultimately turns out to be true, these experiments demonstrate how understanding the mechanisms of visual attention and scene parsing is critical to identifying grammatically specific abilities. Unless one considers attentional factors, one may conclude that infants know a great deal less about grammatical rules and regularities than these children actually do.

5 Conclusions

These experiments demonstrate some of the ways in which interacting factors affect language acquisition. Sensitivity to synchronized auditory-visual information helps in the task of separating speech and thereby aids segmentation. Visual information is often overlooked as a factor in speech perception or regarded as trivial, yet this trivial factor can mean the difference between successful segmentation and an inability to pull out the units of speech in a distracting environment. Similarly in word learning research, we tend to focus strictly on the word learning situation itself, yet prior exposure does matter: Having previously heard a person say a word makes

the difference between successful word learning and failure. Finally, in grammar, research tends to focus exclusively on infants' competence; nonetheless, even a simple attentional distraction can dramatically affect their performance. If the situation is visually very simple, infants can demonstrate knowledge of the meaning of grammatically complex questions. However, if the visual situation is made slightly more complicated, even older infants have difficulty attending to anything other than the object overtly mentioned in a question. The effects in this case were dramatic — a complete and significant reversal. Attentional factors made the difference between successful grammatical understanding and outright misunderstanding.

The common thread in all of these experiments is the importance of multiple interacting factors in language acquisition. A complete explanation of language must encompass not just the specific mechanisms to each task of lexical acquisition but also how speech perception, word learning, and grammar interact. This work provides a demonstration of how methods may be combined to examine such interactions. While future studies by those who do cognitive neuroscience will one day discover the specific neural systems that work to produce these effects, for now, this work serves to reify a truism in brain and language research: Nearly every neural system is involved in the task of acquiring a language. What infants hear affects their visual mappings; what they see affects their segmentation abilities and grammatical understanding. We used to study segmentation and grammar by examining what came into the microphone, yet visual perception matters. We used to study word learning by simply looking at the word learning situation, yet prior exposure matters. With new methods and theoretical advances, we can now look at the interactions between domains as well.

References

ASLIN, R. N., WOODWARD, J. Z., LAMENDOLA, N. P., & BEVER, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Mahwah, NJ: Lawrence Erlbaum Associates.

BAHRICK, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology*, **79**, 253–270.

BAHRICK, L. E., LICKLITER, R., & FLOM, R. (2004). Intersensory redundancy guides infants' selective attention, perceptual and cognitive development. *Current Directions in Psychological Science*, **13**, 99–102.

BALDWIN, D. A., & TOMASELLO, M. (1998). Word learning: A window on early pragmatic understanding. In E. V. Clark (Ed.), *Proceedings of the Stanford Child Language Research Forum* (pp. 3–23). Stanford, CA: Center for the Study of Language and Information.

CAREY, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

CHOMSKY, N. (1957). *Syntactic structures*. New York: Mouton.

DODD, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, **11**, 478–484.

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

ELMAN, J. L., BATES, E. A., JOHNSON, M., KARMILOFF-SMITH, A., PARISI, D., & PLUNKETT, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

FERNALD, A., & KUHL, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, **10**, 279–293.

FERNALD, A., McROBERTS, G., & HERRERA, C. (1992). *Prosodic features and early word recognition*. Paper presented at the International Conference on Infant Studies, Miami, FL.

FORD, D. H., & LERNER, R. M. (1992). *Developmental systems theory: An integrative approach*. Newbury Park, CA: Sage.

HANKIN, J., HOLLICH, G., JUSCZYK, P. W., & SEIDL, A. (2002). *Attention and the comprehension of subject and object* wh-*questions*. Paper presented at the International Conference on Infant Studies, Toronto, Canada.

HOLLICH, G., HIRSH-PASEK, K., & GOLINKOFF, R. M. (1998). Introducing the 3-D intermodal preferential looking paradigm: A new method to answer an age-old question. In C. Rovee-Collier (Ed.), *Advances in infancy research, Vol. 12* (pp. 355–373). NJ: Ablex.

HOLLICH, G. J., HIRSH-PASEK, K., & GOLINKOFF, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, **65**(3), v–123.

HOLLICH, G. J., NEWMAN, R. S., & JUSCZYK, P. W. (2005). Infants' use of visual information to segment speech in noise. *Child Development*, **76**, 598–613.

HOUSTON, D. M., & JUSCZYK, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 1570–1582.

JUSCZYK, P. W., & ASLIN, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, **29**, 1–23.

JUSCZYK, P. W., CUTLER, A., & REDANZ, N. J. (1993). Preference for the predominant stress patterns of English words. *Child Development*, **64**, 675–687.

JUSCZYK, P. W., LUCE, P. A., & CHARLES-LUCE, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, **33**, 630–645.

KUHL, P. K., & MELTZOFF, A. N. (1982). The bimodal perception of speech in infancy. *Science*, **218**, 1138–1141.

LEWKOWICZ, D. J. (1994). The development of intersensory perception in human infants. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of perception: Comparative perspectives* (pp. 165–203). Hillsdale, NJ: Erlbaim.

MASSARO, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge: MIT Press.

MORGAN, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, **33**, 666–689.

NEWMAN, R. S. (2003). Prosodic differences in mothers' speech to toddlers in quiet and noisy environments. *Applied Psycholinguistics*, **24**, 539–560.

NEWMAN, R. S., & JUSCZYK, P. W. (1996). The cocktail party effect in infants. *Perception and Psychophysics*, **58**, 1145–1156.

NOZZA, R. J., ROSSMAN, R. N. F., BOND, L. C., & MILLER, S. L. (1990). Infant speech-sound discrimination in noise. *Journal of the Acoustical Society of America*, **87**, 339–350.

PICKENS, J., FIELD, T., NAWROCKI, T., MARTINEZ, A., SOUTULLO, D., & GONZALEZ, J. (1994). Full-term and preterm infants' perception of face-voice synchrony. *Infant Behavior and Development*, **17**, 447–455.

SAFFRAN, J. R., ASLIN, R. N., & NEWPORT, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928.

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

SEIDL, A., HOLLICH, G., & JUSCZYK, P. W. (2003). Early understanding of subject and object *wh*-questions. *Infancy*, **4**, 423–436.

SMITH, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language* (pp. 277 – 304). Mahwah, NJ: Erlbaum.

THELEN, E., & SMITH, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: The MIT Press.

THORNTON, R. (1995). Referentiality and *wh*-movement in child English: Juvenile d-linkuency. *Language Acquisition*, **4**, 139 – 175.

Van de WEIJER, J. (1998). *Language input for word discovery*. University of Nijmegen, Nijmegen: Wageningen, Ponsen and Loijen.

WAXMAN, S. R., & BOOTH, A. E. (2001). Seeing pink elephants: Fourteen-month-old's interpretations of novel nouns and adjectives. *Cognitive Psychology*, **43**, 217 – 242.

WOODWARD, A. L., & MARKMAN, E. M. (1998). Early word learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology, Volume 2: Cognition, perception, and language* (pp. 371 – 420). New York: John Wiley and Sons.