Early Understanding of Subject and Object Wh-Questions

Amanda Seidl, George Hollich, and Peter W. Jusczyk

Department of Cognitive Science
Johns Hopkins University

In this study, we used the splitscreen preferential looking paradigm to test 13-, 15-, and 20-month-olds’ developing understanding of simple matrix what-questions of the forms “What hit the X?” (subject-question) and “What did the X hit?” (object-question). Infants responded appropriately to subject-questions by 15 months of age, and to both subject- and object-questions by 20 months. At no age did infants look longer toward the object overtly mentioned in the question, as might be expected based on a surface account of early language acquisition. This suggests that infants may have some understanding of these complex structures long before they are produced.

Competence in a language includes competence in syntax. In the last few years, a growing body of research, which relies on data from looking times gleaned from both eye-tracking (Sekerina, Trueswell, & Hill, 1998) and preferential looking studies (Hirsh-Pasek & Golinkoff, 1996; Naigles, 1996; inter alia), has greatly increased our understanding of children’s early abilities to process complex syntactic relations. For example, Hirsh-Pasek and Golinkoff (1996) used the preferential looking paradigm to test 19-month-old infants’ sensitivity to word order in active sentences. They found that, even before infants are able to formulate multi-word utterances, comprehension is guided by word order. Thus, from early on, infants seem to process more than just the individual meanings of...
In addition, younger infants have been shown to exploit both phonological and morphological regularities in language (Gerken, 1996; Jusczyk, Luce, & Charles-Luce, 1994; Mattys & Jusczyk, 2001; Saffran, Aslin, & Newport, 1996; Shi, Werker, & Morgan, 1999; Waxman & Booth, 2001), setting the stage for later syntactic achievements, such as appropriate parsing of wh-questions.

According to linguistic theory, wh-questions are complex structures. For example, for toddlers to comprehend or produce a question such as “what do you think Minnie tried to put on?”, they must both control the relevant syntactic structures and have sufficient processing resources to hold in mind the questioned object for the duration of the sentence. Empirical evidence that these structures are difficult to process comes from production, where children often make errors. For example, Thornton (1995) found that 5-year-old children produce sentences with medial wh-words as shown in (1). Thornton proposes that these errors reflect children’s immature syntax, specifically, problems with the successive cyclic movement of wh-words.

(1) a. Who do you think who’s under there? 
   b. What do you think what Minnie tried to put on?

Thornton (1995)

Given that 5-year-olds make errors in production, one might ask when in development any aspect of these structures is acquired. One possibility is that these structures are not acquired until close to the time when production (even errorful production) begins. Another possibility is that some aspects of these structures are acquired quite early—as are other aspects of syntax (Gerken, 1996; Hirsh-Pasek & Golinkoff, 1996; Naigles, 1996).

Regardless of whether infants actually possess the complex syntactic structures necessary for comprehension, attentional factors would seem to make wh-questions particularly problematic for infants. Consider example (2a). If infants were only to respond to surface features, such as the words present in the sentence, one might expect them to look longer at the object that is overtly mentioned (i.e., apple), rather than the one that provides the answer to the question (i.e., flower). In contrast, where questions such as (2b) have a focus of attention that is also the target, which may make them easier to comprehend. That is, both surface features and deep structure point to the same answer.

(2) a. “What hit the apple?” (When a flower hits an apple)
   b. “Where is the apple?”

Thus, the successful understanding of wh-questions appears to require not only syntactic knowledge, but also the ability to use this knowledge to guide attention even when the correct focus is the object that is not overtly mentioned.

Similarly, in recent work, Lidz and Musolino (2002) show that children’s scopal interpretations are guided primarily by surface syntactic position at a young age.
Given such difficulties, it is reasonable to expect that full comprehension of these questions is a relatively late achievement. Indeed, most previous studies on this topic focused on children who were three years old and older, and who could produce answers to questions (Cairns & Hsu, 1978; Gazdar, 1981; Klima & Bellugi, 1966; Philip, Coopmans, van Atteveldt, & van der Meer, 2000; Radford, 1990).

In this article, we ask whether infants who are not yet readily producing such questions can nevertheless show sensitivity to their form. Because comprehension usually precedes production in language acquisition (Brown, 1973; Gerken, 1996; Hirsh-Pasek & Golinkoff, 1996; Shipley, Smith, & Gleitman, 1969; inter alia), it seems reasonable to ask, what, if anything, children understand about simple matrix wh-questions before the age of two. Our study examines the development of infant’s understanding of questions of the form: “What hit the X?” (called a subject-question because the entity being questioned is the grammatical subject of the verb), and “What did the X hit?” (called an object-question because the entity being questioned is the grammatical object of the verb), following a short video event.

Specifically, given models of development which suggest a change in the weighting from attentional to formal linguistic factors (Hollich, Hirsh-Pasek, & Golinkoff, 2000; Smith, 1999), we asked whether young infants might look at the object that is explicitly mentioned in the question, only later coming to look at the syntactic focus of attention. Furthermore, we asked whether infants would respond appropriately to subject wh-questions earlier than object wh-questions: The object wh-questions are more formally complex and may make more resource demands than the subject wh-questions. Such a subject—object asymmetry in complexity and resource demands has been demonstrated in the adult sentence processing studies (Gibson, 1998; King & Just, 1991), and is also supported by evidence from the processing and acquisition of relative clauses in Korean second language acquisition (O’Grady, Lee, & Choo, 2001; see also Demuth 1992).²³ In addition to their formal complexity, object wh-questions may also be more difficult to process than subject wh-questions because they are less frequent in the input, as has been suggested for object and subject relative clauses (Tabor, Juliano, & Tanenhaus, 1997).

WH-COMPREHENSION IN 13-, 15-, AND 20-MONTH-OLDS

In this Experiment, we tested 13-, 15-, and 20-month-olds’ comprehension of subject what-questions, object what-questions, and where-questions. For both

²For example, in production, either subject- and object-agreement emerge simultaneously, or subject-agreement precedes object-agreement (Demuth, 1992), but it is not the case that object-agreement precedes subject-agreement.

³Because of the structure of Korean, O’Grady et al. (2001) are able to take out the linear distance confound and focus on structural distance. They find that subject relatives are easier than object relatives, even when linear distance should make them more difficult.
what-questions, we showed the infants a video of an event, followed by a what-question that focused attention on one of the objects in a split-screen display. For the where-questions, we simply posed the question while showing infants the split-screen display. We measured whether the infants looked longer at the object that was the answer to the question or the other object on the screen.

Method

Participants. A total of sixty infants were included in the study. The 20-month-old infants had a mean age of 20 months and 14 days (range: 19 months, 26 days to 20 months, 29 days). One additional 20-month-old was excluded from the study due to extreme restlessness. The 15-month-old infants had a mean age of 15 months and 4 days (range: 14 months, 22 days to 15 months, 17 days). The 13-month-old infants had a mean age of 13 months and 4 days (range: 12 months, 24 days to 13 months, 12 days).

Design

The general design is depicted in Table 1. There were five trial blocks: Two tested subject-questions (“What hit the X?”), two tested object-questions (“What did the X hit?”), and one control block tested where-questions (“Where is the X?”). This control was included to establish whether or not the infants could perform when

<table>
<thead>
<tr>
<th>Block</th>
<th>Trial Order</th>
<th>Video</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>Training trial 1 (× 2)</td>
<td>apple hits flower</td>
<td>“What did the apple hit?”</td>
</tr>
<tr>
<td></td>
<td>Test trial 1 (× 2)</td>
<td>split screen of apple &amp; flower</td>
<td></td>
</tr>
<tr>
<td>Block 2</td>
<td>Training trial 2 (× 2)</td>
<td>book hits keys</td>
<td>“What hit the keys?”</td>
</tr>
<tr>
<td></td>
<td>Test trial 2 (× 2)</td>
<td>split screen of book &amp; keys</td>
<td></td>
</tr>
<tr>
<td>Block 3</td>
<td>Training trial 3 (× 2)</td>
<td>apple hits flower</td>
<td>“What hit the flower?”</td>
</tr>
<tr>
<td></td>
<td>Test trial 3 (× 2)</td>
<td>split screen of apple &amp; flower</td>
<td></td>
</tr>
<tr>
<td>Block 4</td>
<td>Training trial 4 (× 2)</td>
<td>book hits keys</td>
<td>“What did the book hit?”</td>
</tr>
<tr>
<td></td>
<td>Test trial 4 (× 2)</td>
<td>split screen of book &amp; keys</td>
<td></td>
</tr>
<tr>
<td>Block 5</td>
<td>Test trial 5</td>
<td>split screen of apple &amp; flower</td>
<td>“Where is the apple?”</td>
</tr>
<tr>
<td></td>
<td>Test trial 6</td>
<td>split screen of book &amp; keys</td>
<td>“Where is the book?”</td>
</tr>
<tr>
<td></td>
<td>Test trial 7</td>
<td>split screen of apple &amp; flower</td>
<td>“Where is the flower?”</td>
</tr>
<tr>
<td></td>
<td>Test trial 8</td>
<td>split screen of book &amp; keys</td>
<td>“Where are the keys?”</td>
</tr>
</tbody>
</table>
the correct answer was the last noun of the question. For the first four blocks, a block began with an action trial (e.g., a book hitting some keys), shown twice in a row (see Figure 1). Then, there were two test trials in which a split-screen display appeared for six seconds, with an accompanying audio track asking either a subject or an object wh-question. In order to reduce memory load and to give infants time to process the question, the subject-, object-, or where-question started one second into the trial, and was repeated at four seconds. There were two such test trials for a total of twelve seconds of testing for each block. The video presentation orders and objects were counterbalanced for the first four blocks. The where-block was always given last.

*Stimulus materials.* The auditory stimuli consisted of recorded versions of subject and object what-questions as well as where questions. The objects used in the experiment (an apple, a flower, a book, and some keys) were chosen because the mapping from word form to word meaning for these objects is reportedly known to infants aged 13-months and older, according to the MacArthur Communicative Development Inventory norms (Fenson et. al., 1994). Examples of the questions used are shown in (3). Observe that (3a) and (3b) are identical except for the slight change in word order and the addition of “did” in (3b). The control where-question (3c) is the only question that overtly mentions the object that answers the question.

![Figure 1](image-url)  
*FIGURE 1* Book hitting keys.
(3) a. Subject-question: “What hit the apple?”
b. Object-question: “What did the apple hit?”
c. Where-question: “Where is the apple?”

These sentences were recorded by the female author, and were read in a happy and child-friendly voice. The durations of the sentences were matched at approximately 1.5 seconds. Digitized versions of these stimuli were saved in a format readable by a Macintosh computer for use in making an audiotrack of the film that the infants heard during testing. During the experiment, the auditory stimuli were presented at 72dB.

Objects were constructed using a 3D drawing program. They were brightly colored and large, so as to attract the infants’ attention, yet designed to be of approximately equal salience. The resulting pictures were turned into action trials, in which they performed a hitting/bumping action (Figure 1). The actions were created using a digital video construction program (EditDV produced by Digital Origin). Split-screen displays were created in which the two objects appeared side-by-side (Figure 2). The individual pictures were 18 in × 12 in, and were placed 20 in. apart from center-to-center. (A sample of the video is available at http://www.

4 Both the agent of the action and the patient were affected by the hitting action so that an infant’s attention would not focus only on the object that moved.
infancyarchives.com.) The same program was then used to combine the audio and video into the counterbalanced orders that were finally shown to the participants.

Apparatus

A Sony TRV-7000 Digital8 camcorder was attached to a Sony KLV-9000 LCD presentation display (a 56-in. screen). A second camcorder was mounted above this display to record infant looking times. A large white plywood barrier surrounded the front of the display and camera in order to minimize looks to non-essential sections of the testing apparatus. In this way, the infant or toddler was prevented from seeing anything except the screen and the lens of the camera.

Procedure

Upon arrival at the lab, parents and their infants were escorted to a playroom. There, the experimenter explained the study and gave the parent a consent form. The infant and parent were then led to a testing room, where the infant was seated in the center of their parent’s lap and the parent was directed not to interact with the infant during the experiment. During the experiment the parent wore a visor with a long piece of felt hanging from the front. This blocked the parent’s view of the screen and prevented the parent from influencing the child to look in any particular direction. The parent and child were seated 46 in. from the screen. The entire experiment took 2 min and 16 sec.

Coding

For each subject, durations of looking time to the left or the right were coded off line, frame-by-frame, and without sound in order to minimize the possibility of bias. The first author coded all the data. The second author coded ten percent of the data. Inter-coder reliability was .96 or better (this is the percent of agreement between coders).

Results

Table 2 shows mean looking times in seconds to the target and the non-target for the three question types across all the age-groups, as well as the number of infants whose means are positive (meaning they looked longer toward the target during the test trial). The data show different patterns for the different age groups. Planned comparisons for the 20-month-olds showed that there were significant
differences in target and non-target looking times for the subject, \( t(19) = 3.482, p = .0025 \), the object-, \( t(19) = 5.256, p < .0001 \), and the where-questions-, \( t(19) = 5.950, p < .0001 \). Fifteen-month-olds showed reliably longer looking times to the target for the subject-, \( t(19) = 2.218, p = .0389 \) and the where questions \( t(19) = 2.576, p = .0185 \), but not for the object-questions \( t(19) = .721 \). For the 13-month-olds, none of the three comparisons revealed statistical significance (subject- \( t(19) = 1.014 \) ns, object- \( t(19) = –.101 \) ns, or where questions \( t(19) = .904 \) ns).

These findings indicate that 20-month-olds are able to respond appropriately to simple subject-, object-, and where-questions, and that 15-month-olds do so for simple subject- and where- questions, but not for object-questions.\(^5\) The 13-month-olds were unable to respond appropriately to any of the question types. It was especially surprising that the 13-month-olds were unable to respond appropriately to the simple where-questions. Because of this result, we ran an additional experiment with only the where-questions. In this experiment, we exposed infants to the actions immediately before each where-question, and then repeated each question four times during each test trial. Nonetheless, the results were the same as in Experiment 1: The infants did not respond above chance.\(^6\) Given the results of this control study, it appears that our task may have been too difficult for the youngest infants, i.e., the attentional demands of processing the moving stimuli and quickly shifting to a question, proved too difficult for infants at this age. Indeed, given other studies that

\(^5\)A time-course analysis of these results revealed that a substantial portion of infants shifted their looking to the target object approximately 1 sec after the onset of the critical word; well after the sentence had ended. This suggests that it was the sentence as a whole that was driving their looking times, rather than any particular word per se.

\(^6\)We also ran individual item analyses in order to ascertain whether the infants were simply unable to understand the target nouns in the questions. However, again, a null result was found for each individual item. We therefore conclude that the difficulty with all of the questions for the 13-month-olds may have more to do with attention than specific vocabulary.
suggest infant word learning is extremely tenuous at this age (Werker, Fennell, Corcoran, & Stager, 2002), perhaps it should not come as a surprise that attentional distraction may disrupt competence in this case as well. Thus, in the subsequent analyses, we consider only the data from the 15- and 20-month-olds. Nonetheless, the 13-month-old data is reported, because, although the 13-month-olds did not significantly look at the object overtly mentioned in the subject questions, their data does fall inline with the results of the older age groups.

Planned comparisons on 15- and 20-month-olds’ difference scores show a developmental trend for object-questions, $F(1, 38) = 12.12, p = .0013$, but not for subject-questions, $F(1, 38) = .358, p = .5533$. These results suggest that, although comprehension of all questions improves and looking times to targets increase with age, comprehension of object-questions shifts between 15 and 20 months.7

**DISCUSSION**

The results of this study indicate that by 20 months of age, infants are sensitive to simple matrix what-questions of both subjects and objects. Infants correctly looked at the target object when asked both “What hit the flower?” (the apple) and “What did the apple hit?” (the flower). This finding is significant, given findings from previous studies suggesting that comprehension of such questions is quite a late achievement. For example, some of the youngest children (three-year-olds) previously tested in studies by Ervin-Tripp (1970) and Tyack & Ingram (1977) were unable to correctly answer either subject- (Ervin-Tripp, 1970) or object- (Tyack & Ingram, 1977) questions. However, Stromswold (1995), in her review of the literature on acquisition and the comprehension of wh-questions, mentions that the low success rate of these studies may be related to the high task-related demands of these studies. Consistent with Stromswold’s discussion, it may be that the infants in these earlier studies did, in fact, understand some aspects of the questions they were asked. However, requiring children to produce a verbal response masked their understanding. In contrast, the low task demands in our study reveal that syntactic competence may begin at an earlier age than might have been previously expected.

Aside from the 20-month-olds’ overall competence, we also observed a shift in performance between 15- and 20-months. Only the 20-month-olds showed understanding of both the object and the subject what-questions. This suggests that there may be a developmental change that is a prerequisite for understanding of object-questions. What would account for this developmental change? We consider three alternatives.

One possibility is that infants parse language based simply on surface features or semantic features. Namely, when young infants hear a question word such as

---

7We also analyzed these data by the longest look, instead of overall looking time. Because these analyses overlapped completely with the overall results, we do not include them here.
“what” or “where,” they either look at the word overtly mentioned in the sentence containing the wh-word or focus on the agent of the action. This finding would be predicted by accounts that argue that early on infants rely on attentional, social/pragmatic, or semantic foci to process language (Baldwin, Markman, Bill, Desjardins, & Irwin, 1996; Pinker, 1984; Smith, 1999; Tomasello, 1999). We can rule out this first alternative for 15- and 20-month-olds. The fact that children respond appropriately to what-questions eliminates the possibility that they are merely looking at an object overtly mentioned in the target sentence. In fact, in these questions, they look longest at an object that is not mentioned in the sentence. It is also true that toddlers do not just focus on an agent or patient of some action, because by 20 months of age they were able to respond appropriately regardless of which object (agent or patient) was being questioned, although an agentive preference could explain the performance of the 15-month-olds.

This result has implications for a central debate in the language acquisition studies that focus on whether infants begin the task of acquisition by exploiting linking rules from thematic roles (agent/patient) observable in scenes to syntactic categories (subject/object) (Grimshaw, 1981; Pinker, 1984, 1994) or whether infants rely on information in the syntactic frame together with other information in the scene to infer the meaning of a verb (Brown, 1957; Gleitman, 1990; Landau & Gleitman, 1985). We have found that, even if infants begin with a preference for agent, by 20 months of age they are able to override this preference and rely primarily on syntactic information to answer these questions.

In addition, it is clear that the strategy of focusing on the noun which is not overtly mentioned in the sentence is unique to these types of questions, because in where-questions, infants looked at the item that was overtly mentioned in the sentence. At the very least, our data show that infants distinguish what- and where-questions. This is particularly interesting given that the two sentences are phonologically similar (e.g., “Where is the flower?” vs. “What hit the flower?”).8

A second possibility is that 20-month-old infants have control of complex syntactic structures, but 15-month-olds do not have full control of these structures [either do-support (insertion of the dummy verb do) or movement chains (movement of the question word what to its surface position in the sentence)].9 It may be

---

8It is also possible that the 15-month-olds had an easier time with both subject- and where-questions (as opposed object-questions) because these questions end with nouns. These nouns may have heightened 15-month-olds overall attention to the meaning of the utterance, in the case of subject-questions. Indeed, work by Golinkoff, Alioto, and Hirsh-Pasek (1995; see also Fernald, McRoberts, & Herrera, in press) suggests that infants find it easier to recognize words presented in sentence final position. While current studies are addressing this issue, certainly by 20 months, infants have overcome this limitation.

9Note that the problems we suggest 15-month-olds may have with object-questions are problems of processing, not of the grammatical system. Syntactically, objects have fewer restrictions on extraction than subjects, but this does not necessarily contribute to helping, or hindering, the processing of sentences with extracted objects.
that the 15-month-old infants are thus forced to use a simpler strategy based on semantic roles, rather than the deeper syntactic strategy exploited by the 20-month-old infants, and that exploiting this strategy rather than the syntactic one may lead them astray.\textsuperscript{10} The syntactic strategy, according to a standard theory of wh-movement, is that the wh-word moves to the specifier position of the complementizer phrase (Chomsky, 1981). Given this view, object-questions may be acquired later than subject- and where-questions for several reasons.\textsuperscript{11} First, do-support may make object-questions harder than subject-questions.\textsuperscript{12} Second, object-questions may be harder to understand than subject ones, because they involve longer movement.\textsuperscript{13} This is true regardless of if one favors a theory that involves successive cyclic movement (Chomsky, 1995) or one such as \textit{Tree Adjoining Grammar} (Joshi, 1986), which involves a long chain between a wh-trace and its surface position. In that case, the object needs to move a bit farther than the subject, which may mean more working memory is involved.

This brings us to a third possibility: 15-month-olds’ problems with object-questions may result from a lack of resources either in terms of working memory or central executive function. In a similar way, it has been argued that patients with Parkinson’s disease (patients who have problems with working memory and central executive function) have resultant problems with complex sentence

\textsuperscript{10}Something similar is suggested in Hirsh-Pasek and Golinkoff (1996): Two-year-old boys treated sentences such as, “Big Bird is squatting with Cookie Monster,” as a transitive sentence: Girls appeared to use syntactic cues to correctly interpret these sentences as intransitive sentences with conjoined subjects.

\textsuperscript{11}Although children may not have higher projections such as CP in their production (as suggested in Legendre et al., 1999), it does not necessarily follow that they cannot comprehend sentences with higher projections at an earlier age.

\textsuperscript{12}In a generative account of English, do-support occurs because lexical verbs do not raise. Thus, in order to check tense features that occupy C (a complementizer-head above TP), the formal features of a verb may move to C. In object-questions, when a subject is present in Spec vP (the specifier position of a phrase headed by voice, or in an account without vP, Spec VP) it blocks the feature checking relationship between C and the lexical verb, cf (5) and (4). But these features must be checked for the derivation to go through, thus do is inserted as a last-resort operation and checks off the features in C and T (Chomsky, 1995).

\begin{align*}
(4) \text{[what[c[t[have]]v[hit[flower]][vP][iP][TP]][C][CP]} \\
(5) \text{[what[c[t[had][apple][v[hit[t[NP][vP][iP][TP]][C][CP}]
\end{align*}

\textsuperscript{13}An anonymous reviewer points out that our argument about the length of the movement chain does not explain why it is that children understand where-questions before object-questions, since both involve long movement chains. Although the chain may be equally long in where- and object-questions, the movement is of an adjunct, not an argument, in where-questions, and thus we do not necessarily predict the same pattern of results. In addition, in where-questions, it is possible to understand the target in the absence of processing it as a question at all. Nonetheless, it could be that a combination of factors (both the length of the chain and the fact that the entity being questioned is not overtly mentioned) is what makes object-questions especially difficult.
processing (Caplan & Waters, 1999; Grossman et al., 1992; Natsopulos et al., 1991). It has also been shown that normal adults have more difficulty in processing object-relative clauses when they have the additional resource demands of a secondary task (Gibson, 1998; King & Just, 1991). Similarly, Santelmann & Jusczyk (1997) found changes in infants’ processing between 15 and 18 months of age. Thus, it may be that infants have complete control of the structures necessary to process both object and subject wh-questions, but, at 15 months, do not have sufficient working memory to process questions that require more resources. Ongoing research is examining the degree to which attentional demands may cause infants to have difficulty in processing more complex structures.

Regardless of which explanation ultimately turns out to be correct, this study indicates that, long before infants are producing wh-questions, they understand them in ways few would have suspected.

ACKNOWLEDGMENTS

This research was supported by an NICHD research grant (#15795) and by a NIMH Senior Scientist award (#01490) to Peter W. Jusczyk, and by an NSF IGERT postdoctoral fellowship to Amanda Seidl. The authors are grateful to Anne Marie Jusczyk and Natasha Henline for help in recruiting infants. We would also like to thank Barbara Landau, three anonymous reviewers at Infancy, and Leslie Cohen for their comments and suggestions on this work. A portion of this work was previously published in the Proceedings from BUCLD 26. This work is dedicated to Peter Jusczyk, our fellow collaborator, mentor, and friend.

REFERENCES


