U-shaped development in the acquisition of filler-gap dependencies: Evidence from

15- and 20-month olds

Annie Gagliardi, Tara Mease & Jeffrey Lidz

Department of Linguistics

University of Maryland, College Park

1401 Marie Mount Hall
University of Maryland
College Park, MD 20742-7505
Tel: +1-301-405-8220
Fax: +1-301-405-7104

acg39@umd.edu
jlidz@umd.edu
Abstract

This paper investigates infant comprehension of filler-gap dependencies. Two experiments probe 15- and 20-month-olds’ comprehension of two filler-gap dependencies: *wh*-questions and relative clauses. Experiment 1 shows that both age groups appear to comprehend *wh*-questions. Experiment 2 shows that only the younger infants appear to comprehend relative clauses. We argue that this surprising U-shaped pattern follows from an offset in the development of grammatical knowledge and the deployment mechanisms for using that knowledge in real time. 15-month-olds, we argue, lack the grammatical representation of filler-gap dependencies but are able to achieve correct performance in the task by using a parsing heuristic based on argument structure. 20-month-olds, we argue, do represent filler-gap dependencies, but are inefficient in deploying those representations in real time.

*Keywords*: infant language comprehension; filler-gap dependencies; U-shaped learning
U-shaped development in the acquisition of filler-gap dependencies: Evidence from 15- and 20-month olds

1 Introduction

Human language abilities can be broken down into two main components: the knowledge of the linguistic system and the deployment of this knowledge. While studies of adult psycholinguistics have primarily focused on the deployment of this knowledge, holding grammatical knowledge constant, studies of child psycholinguistics have primarily focused on when different aspects of linguistic knowledge are learned, and what information is available to aid the child in learning. This approach, however, only taps into a part of what governs a child’s linguistic behavior.

We can think of language acquisition as the fitting together of the two components: the child must not only acquire the linguistic knowledge, typically viewed as a kind of declarative knowledge (Chomsky 1965), but also must learn to tie this knowledge to an appropriate deployment system that enables speech production and comprehension in real time (Frazier & Devilliers 1990, Chang, Dell & Bock 2006). In order to investigate the relation between the acquisition of grammatical knowledge and the accompanying deployment system, it will be useful to find two cases where we believe the underlying knowledge to be similar, but the deployment processes to differ based on surface properties. We find such a distinction in the processing of two filler-gap dependencies: wh-questions and relative clauses. Due to both their unbounded nature and uniquely
linguistic character, filler-gap dependencies are an ideal place to examine the relation between grammatical knowledge and deployment in adult processing and language development. In our investigation of the acquisition of these two types of dependencies, we observe a case of U-shaped development that can be explained in terms of growth of grammatical knowledge with a delay in the real-time deployment mechanisms.

This paper proceeds as follows. Section 2 reviews (a) the linguistic evidence supporting the view that *wh*-questions and relative clauses access the same linguistic knowledge, (b) the psycholinguistic models of how this knowledge is deployed in real time and, (c) the current understanding of the developmental time course of these dependencies. In Section 3 we describe a set of experiments revealing that both 15- and 20-month-old infants appear to understand *wh*-questions. In Section 4 another set of experiments shows that 15-month-olds, but not 20-month-olds, appear to understand relative clauses. Section 5 will be a discussion of these surprising results in light of the relationship between knowledge and deployment in language acquisition, making a case for a nonadultlike parsing heuristic in younger infants that is replaced by adultlike mechanisms in older ones.

2 Background: Filler-Gap Dependencies

Filler-gap dependencies are a class of dependencies in human languages that relate an element in a non-thematic position (henceforth the 'filler', shown in italics) to its
canonical thematic position in the sentence (henceforth the 'gap', marked by ___).¹ These dependencies can be quite local (1) or arbitrarily long (2).

(1) Which dog did the cat bump ___?
(2) Which dog did the monkey think that the horse saw the cat bump ___?

Among wh-questions, there are two surface form differences between extractions from subject positions and object positions. First, displacement is farther and therefore more apparent for object extraction (3) than for subject extraction (4). Second, within a single clause subject questions do not require subject-object inversion (3) but object questions do (4).

(3) Which dog ___ bumped the cat?
(4) Which dog did the cat bump ___?

However, it is widely agreed that the set of grammatical mechanisms responsible for generating subject extraction is the same as that generating object extraction. Evidence for this lies in the fact that both types of displaced elements can be related to their thematic positions across finite clauses, as in (5)-(6), that both types are sensitive to island constraints (7-8) and induce island effects for other dependencies (9-10) (Ross 1967; Chomsky 1986, Rizzi 1990).

¹ We restrict our attention in this paper to dependencies involving arguments.
Which dog do you think ___ bumped the cat?

Which dog do you think the cat bumped ___?

* Which dog did the man make the claim that ___ bumped the cat?

* Which dog did the man make the claim that the cat bumped ___?

* How did the man wonder [which dog, ___ bumped the cat ___]?

* How did the man wonder [which dog, the cat bumped ___]?

* Wh-questions (1) are only one type of filler-gap dependency. Another structure, the relative clause (11) is also a filler-gap dependency.

Show me the dog that the cat bumped ___

There are several differences in the surface properties of wh-questions (1) and relatives (11): the presence of a wh-word in (1) and its absence in (11); the fact that the filler is always clause initial in a wh-question but not in a relative clause; the lack of subject-auxiliary inversion in relative clauses; and, when uttered aloud, prosodic differences between the two sentences. Despite these differences, however, we have reason to believe that the same grammatical mechanisms are at work in their generation. Both involve the displacement of the filler from its thematic position to a higher position. The displacements appear to be parallel, as the fillers in both dependency types are unbounded, but can only originate in certain, parallel, structural positions (12)-(14) (Chomsky 1977).
(12) a. Which dog did you think (that she said) the cat bumped __? 
   b. Show me the dog that you thought (that she said) the cat bumped __?

(13) a. *Which dog did the monkey think that ___ bumped the cat?  
   b. *Show me the dog that the monkey thought that ___ bumped the cat

(14) a. *Which dog did the cat bump the monkey and ___?  
   b. *Show me the dog that the cat bumped the monkey and ___

The comprehension of both types of filler-gap dependencies is also expected to be driven by a similar mechanism, as both dependencies require the comprehender to somehow link up the filler with the gap. Below is an overview of the process thought to be responsible for the resolution of filler-gap dependencies by adults, and of early knowledge of these dependencies.

2.1 Adult Parsing

It is widely agreed that adult speakers resolve filler-gap dependencies using an active filling strategy (Crain & Fodor 1985, Frazier & Clifton 1989, Frazier & Flores D'Arcais 1989, Traxler & Pickering 1996, Sussman & Sedivy 2003, Aoshima et al 2004). In an active filling strategy, as soon as a filler is encountered, the search for a potential gap site begins. Comprehenders could identify a filler because of its displacement from its canonical position in the sentence, the intonation contour of the utterance and other features such as wh-words and scope markers. Gap sites would be posited at every structural position where an argument could occur. Convergent crosslinguistic evidence
for this strategy comes from both reading time and ERP measures, which find a disturbance when the first potential gap site encountered by the parser is already filled (15) or when it is not the predicted position based on semantic information found in the filler (16) (Stowe 1986; Traxler et al 2002).

(15) My brother wanted to know who Ruth will bring us home to at Christmas

(16) The scientist that the climate annoyed ___ did not interest the reporter

Active filling is not the only possible strategy for resolving filler-gap dependencies, however. Another strategy that parsers might engage would be gap driven parsing (Wanner & Maratsos 1978). In gap driven parsing, the parser begins a backwards search for a filler only when it encounters the gap site. While there is ample evidence against gap driven parsing in adults (Frazier & Flores D'Arcais 1989, Traxler & Pickering 1996, Sussman & Sedivy 2003, Aoshima et al 2004), it is worth mentioning as a potentially plausible strategy, especially when considering the development of filler-gap parsing in children.

The processing of all filler-gap dependencies does not seem to be equal, however, and various researchers have found that subject gaps (17) are easier to resolve than object gaps (18) (Gibson 1998).

(17) Show me the dog ___ that bumped the cat

(18) Show me the dog that the cat bumped ___
This asymmetry (indexed by slower reading times and poorer comprehension of object gaps), is not absolute, and can be modulated by factors including working memory load, animacy of arguments, plausibility of predicates, distance of extraction and the amount and type of intervening material (Konieczny 2000, Gordon et al 2001, Traxler et al 2002, Mak et al 2002, Fiebach et al 2002, Clifton et al 2003). The study of the subject-object asymmetry has focused on long distance (multiclausal) extractions, and has mainly looked at the processing and comprehension of relative clauses. Asymmetries like this one are evidence of the apparent disjunct between knowledge and deployment. Whereas the grammatical mechanisms for characterizing subject and object dependencies are similar, the deployment, or real time resolution of the dependencies, reveal differences.

While the subject-object asymmetry has been deeply investigated, few studies directly compare the processing of wh-questions and relative clauses. Based on the superficial differences between the constructions mentioned above, it is possible there is an asymmetry between them in online parsing.

2.2 Acquisition of Filler-Gap Dependencies

Various researchers have looked at the acquisition of wh-questions and relative clauses. In particular, the first productions of these constructions have been studied, both by looking at naturalistic child utterances from transcripts, and by eliciting relative clauses and wh-questions (Hamburger & Crain 1982, deVilliers et al 1990, Stromswold 1995,
Thornton 1995). Early comprehension of relative clauses has mainly been studied by act-out tasks (Tavakolian 1981, Hamburger & Crain 1982), and early comprehension of wh-questions by question answering tasks (Roeper & deVilliers 1994, deVilliers & Roeper 1995, Goodluck in press). These studies have focused on finding out when children are able to properly deploy their knowledge of filler-gap dependencies, and has looked at whether surface form differences found within a dependency type (i.e. subject vs object extraction) affect the age of acquisition. While individual findings vary, there does not appear to be straightforward evidence either for or against a subject-object asymmetry in the order of acquisition of filler-gap dependencies. What is clear is that from as young as can be tested children appear to follow adult-like constraints of the formation of filler-gap dependencies, effectively deploying their knowledge of these constructions. While the acquisition of both relative clauses and wh-questions has been studied, no studies have drawn direct comparisons between the dependency types, and it is thus unclear how parallel the acquisition of these two types of dependency is. Importantly, all of these studies looked at the acquisition of filler-gap dependencies once children were producing them. As we generally find that production lags behind comprehension in development, it is likely that children are able to deploy their knowledge of these dependencies for comprehension earlier than for production.

Only one study that we know of has looked at the pre-production comprehension of filler-gap dependencies. Seidl, Hollich and Jucszyk (2003) used the intermodal preferential looking procedure to examine comprehension of wh-questions by 13-, 15- and 20-month-olds. Each infant was tested on the comprehension of two subject questions, two object
questions and one *where* question. They found that 20-months olds appeared to understand all three question types, 15-months olds appeared to understand only subject and *where* questions, and 13-month-olds did not appear to understand any question type. They suggested that the subject-object asymmetry found in the 15-month-olds was due to either the longer structural distance between the filler and the gap in object questions as compared with subject questions, or the fact that the infants were not yet equipped to deal with the *do*-support employed in object questions. Exploring whether the 15-month-olds' failure at object questions reflects a lack of grammatical knowledge or an inability to properly deploy this knowledge lies behind the motivation for the current experiments.

3 Experiment 1: *WH*-questions

3.1 Motivation

Determining whether a lack of knowledge or an inability to deploy knowledge lies behind the 15-month-olds' reported difficulty with object questions is the first step in investigating the mechanisms behind the development of the parsing of filler-gap dependencies. To do so, we first need to take a closer look at the Seidl et al study. While Seidl et al cited the longer structural distance and *do* support as the two factors which could have made object extraction too difficult for 15-month-olds, the situation is in fact more complex. Possible explanations of 15-month-olds' poor performance on object-questions can be roughly broken into two linguistic hypotheses, the *Structural Distance Hypothesis* and the *Do-Support* hypothesis, and one methodological hypothesis, the
Methodological Hypothesis. The linguistic hypotheses can each in turn be broken down into hypotheses regarding knowledge and deployment.

The Structural Distance Hypothesis posits that the longer distance between the filler and the gap in object questions causes the 15-month-olds' difficulty. This difficulty could derive from the infant lacking the grammatical knowledge needed to compute displacement, which is necessary in object questions but could be viewed as optional in subject questions, as the position of the subject is identical in monoclausal declaratives and monoclausal wh-questions (George 1980, Chung & McCloskey 1983). Alternatively, the child might possess this knowledge but be unable to deploy it effectively when the filler is far away from the gap, as in object questions (Gibson 1998).

The Do-Support Hypothesis posits that do-support is responsible for the difficulty. This difficulty could derive from the child lacking the requisite knowledge of functional structure that is needed to interpret do-support (e.g., Radford 1990). Alternatively, there could be a parsing problem when this knowledge is deployed. For example, if do is misanalyzed as a main verb the remainder of the parse, and associated comprehension processes would be disrupted.

The Methodological Hypothesis predicts that factors in the design and materials employed by Seidl et al could have masked the infants’ underlying linguistic abilities. As mentioned above, each infant saw two trials of each question type in a within subjects design. Two trials per questions type may not have given infants sufficient time to adjust
to task demands, and the within subjects design may have caused interference between the two question types. Additionally, the stimuli consisted of two-dimensional cartoons of two inanimate objects floating through space and colliding, followed by a test phase where the two objects were presented side by side along with wh-question audio. This type of animation was unengaging and also pragmatically odd. Because only one event took place, the question was pragmatically infelicitous. Only one thing could possibly be the answer.

In the first experiment we set out to determine whether the asymmetry seen in the 15-month-olds in the Seidl et al study was due to one of the linguistic hypotheses or the methodological one. In order to investigate these hypotheses and identify the source of 15-month-olds’ difficulty with object questions, we made several manipulations to the basic design of the Seidl et al. study. Target utterances were wh-questions patterned after those in (19):

(19)  Subject WH Question: Which dog bumped the cat?
Object WH Question: Which dog did the cat bump?

To probe the methodological hypothesis we attempted to improve upon the factors we identified as potentially problematic above. First, we employed a between subjects measure, allowing for six trials per subject, all of the same question type. This would give the infants ample time to adjust to the task and eliminate the potential interference of question type. Employing six trials also allowed us to analyze the data by blocks,
enabling us to determine whether having too few trials can hide children’s knowledge. To improve the stimuli, we used videos of engaging puppets, with three characters per scene. The addition of an extra character served two functions. First, it made the question felicitous. If two animals separately performed the same kind of action, it is plausible that a speaker might be unsure of who did what to who, motivating the use of a question. Additionally, the third character provided the felicity conditions necessary for a relative clause, i.e. the differentiation between two different dogs requires the sort of information specifiable in a relative clause.

3.2 Predictions

The predictions for this first experiment are straightforward. Regarding 15-month-olds, if the Methodological concerns were responsible for the 15-month olds’ asymmetry in the Seidl et al experiment, then these asymmetries should disappear when these concerns have been addressed. In addition, if these methodological concerns are compounded by the use of too few trials, then we predict an effect of block, with 15-month-olds showing greater success in later trials than in early trials. If either the Do-Support or Structural Distance hypotheses were behind the asymmetry, it should maintain. 20-month-olds are predicted to behave the same way as they did in the Seidl et al study.

3.3 Participants
32 15-month-olds (17 males) with a mean age of 15;0 (range: 14;14 to 15;18) and 32 20-month-olds (17 males) with a mean age of 20;03 (range: 19;07 to 20;22) were included in the final sample. Participants were recruited from the greater College Park, MD area and were acquiring English as native language. Parents completed the MacArthur Short Form Vocabulary Checklist. 15-month-olds' mean production CDI-vocabulary was (19.2) (range: 0 to 60), and 20-month-olds’ mean production CDI-vocabulary was (125) (range: 21 to 574). We analyzed the data of infants that completed at least 4 out of 6 test trials, and the trials where the infant was looking at least 20% of the time. Nine additional infants were tested but ultimately excluded from the analysis due to fussiness or inattention.

3.4 Materials

3.4.1 Visual Stimuli

We first created digitalized video recordings of puppets performing the actions on one another. This footage was edited to create the series of events outlined in Table 1 below. All sequences were filmed against a white background and presented on a 51” plasma television screen. A sample video of an entire trial can be found at (http://www.ling.umd.edu/labs/acquisition/stimuli/wh_s_bump).

3.4.2 Auditory Stimuli
The audio portion of the stimuli (as outlined below in Table 1) were recorded in a soundproof room by a female speaker of American English in an infant friendly voice. These recordings were edited and combined with the visual stimuli. For consistency, wherever the audio was identical across trials, the same recording was used.

### 3.5 Apparatus and Procedure

Each infant arrived with his/her parent and was entertained by a researcher with toys while another researcher explained the experiment to the parent and obtained informed consent. The infant and parent were then escorted into a sound proof room, where the infant was either seated on the parent's lap or in a high chair, centered six feet from a 51" television, where the stimuli were presented at the infant's eye-level. If the infants were on the parents' laps, the parents wore visors to keep them from seeing what was on the screen. Each infant was shown six trials, all from the same experimental condition. Each experiment lasted 6 minutes, and the infants were given a break if they were too restless or started crying. The infant was recorded during the entire experiment using a digital camcorder centered over the screen. A researcher watched the entire trial with the audio off on a monitor in an adjacent room and was able to control the camcorder’s pan and zoom in order to keep the infants face in focus throughout the trial.

The procedure included three phases: character familiarization, action familiarization and a test phase (See Table 1). Each trial consisted of these three phases, and each infant watched six trials. Each trial consisted of a different combination of animals and action
(e.g., two dogs, a cat and a bumping action; two mice, a bee and a tickling action). (See Appendix A for complete descriptions). To focus infants' attention before the beginning of each trial, a four second still of a smiling infant, combined with an audio track of an infant giggling, was shown. Trials were presented in one of two random orders, balanced across conditions. The direction of the action (right to left or left to right) was counterbalanced across the orders. The screen position of the characters was kept constant from action familiarization to test, and the left-right position of the target animal was counterbalanced across conditions. Infants were randomly assigned to the one of two orders in the WH-subject or WH-object condition. Infants saw the exact same videos across conditions, with only the audio portion varying.

3.5.1 Character Familiarization Phase

(20 sec) Infants were introduced to each of the animals that would be involved in the action (4s each, followed by a 1s black screen break), and then shown a shot of the three animals together (also 4s). The accompanying audio varied as a function of both trial and condition. For example, a white dog was introduced and the infants heard, "Hey look! It's a white dog". This was followed by similar introductions of a brown dog and a cat. When the white dog, the cat and the brown dog were all together, the infant heard, "Somebody's gonna bump the cat" (subject condition) or "The cat's gonna bump somebody" (object condition). The characters were always arranged with the single animal in the middle, flanked by the animals of the same species (e.g. white dog - cat - brown dog).
3.5.2 Action Familiarization Phase

(17 sec) Infants saw a clip containing a series of two actions, followed by a black screen break, followed by the same video clip. In each scene the animal on the far left or right (e.g. the white dog) would perform an action (e.g. bumping) on the middle animal (e.g. the cat), who in turn performed that same action on the animal on its other side (e.g. the brown dog). During the first video clip, the infants heard the attention direction audio "Look what's happening! Do you see it? Wow!". During the black screen break the infants heard audio that varied by condition, e.g. "Which dog is gonna bump the cat?" (subject condition) or "Which dog is the cat gonna bump? (object condition).

3.5.3 Test Phase

(15.3 sec) During the test phase the infants were presented with the two animals of the same kind (e.g. the two dogs), one on either side of the screen, consistent with their position during the action phase. After 0.6 seconds the infants heard "Now look!", followed by the target question, which varied as a function of condition (e.g. "Which dog bumped the cat?", subject condition). This presentation lasted 6 seconds and was followed by a black screen for 3.3 seconds, during which the target question was repeated. The offset of the target question was aligned with the presentation of the two animals once again. One second later the infants heard "Can you find him?" followed by a reiteration of the target question.
### Table 1

<table>
<thead>
<tr>
<th>Number of Frames</th>
<th>Video</th>
<th>Audio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00</td>
<td>Black Screen</td>
<td>none</td>
</tr>
<tr>
<td>4:20</td>
<td>Smiling Baby</td>
<td>4:00 Baby Giggle</td>
</tr>
<tr>
<td>1:00</td>
<td>Black Screen</td>
<td>none</td>
</tr>
<tr>
<td>4:00</td>
<td>White dog</td>
<td>Hey look! It’s a white dog</td>
</tr>
<tr>
<td>1:00</td>
<td>Black Screen</td>
<td>none</td>
</tr>
<tr>
<td>4:00</td>
<td>Brown dog</td>
<td>Now look! It’s a brown dog</td>
</tr>
<tr>
<td>1:00</td>
<td>Black Screen</td>
<td>none</td>
</tr>
<tr>
<td>4:00</td>
<td>Cat</td>
<td>Now look! It’s a cat</td>
</tr>
<tr>
<td>1:00</td>
<td>Black Screen</td>
<td>none</td>
</tr>
<tr>
<td>4:00</td>
<td>All animals</td>
<td>Somebody’s gonna bump the cat*</td>
</tr>
<tr>
<td>1:00</td>
<td>Black Screen</td>
<td>none</td>
</tr>
<tr>
<td>7:00</td>
<td>White dog bumps cat, Cat bumps brown dog</td>
<td>Look what’s happening! Do you see it? Wow!</td>
</tr>
<tr>
<td>3:00</td>
<td>Black Screen</td>
<td>Which dog’s gonna bump the cat?*</td>
</tr>
<tr>
<td>7:00</td>
<td>White dog bumps cat, Cat bumps brown dog</td>
<td>Look what’s happening! Do you see it? Wow!</td>
</tr>
<tr>
<td>1:00</td>
<td>Black Screen</td>
<td>none</td>
</tr>
<tr>
<td>6:00</td>
<td>Split Screen: White dog, Brown dog</td>
<td>Now look! Which dog bumped the cat?*</td>
</tr>
<tr>
<td>3:10</td>
<td>Black Screen</td>
<td>Which dog bumped the cat?*</td>
</tr>
<tr>
<td>6:00</td>
<td>Split Screen: White dog, Brown dog</td>
<td>Can you find him? Which dog bumped the cat?*</td>
</tr>
</tbody>
</table>

Audio segments marked by an asterisk (*) varied as a function of condition.

### 3.6 Coding

The event and character portions of the videotaped sessions were coded off-line to track infants’ attentiveness to the familiarizations. Test portions of the video sessions were also coded off line. The sound was turned off and coders were blind as to which condition the videos were from. Using Supercoder (Hollich 2003) coders went through the videos
frame by frame (29.97 frames per second) and noted whether the infant's gaze was directed to the left or right of the screen, or if they were looking away. Collecting frame by frame results for each infant's looking patterns in every trial we were then able to analyze the data in two ways.

First, in each condition we were able to compile the total proportion of looks toward the target animal for each frame. Combining these proportions gave us a timeline of proportion of looks towards the target for every frame in the test trial. This time line allowed us to look for general trends in looking across the trials.

We were also able to analyze particular critical time-windows, by averaging the proportion of participants looking towards the target for a certain number of frames. We used this method to look at the average proportion of looks towards the target animal in a one second baseline before the target question was uttered, and similarly for windows following each iteration of the target question. It is the averages that we found in these target windows, that we will be comparing below.

Four coders coded this data. Inter-coder reliability was always above 90% and Cohen’s Kappa ≥ 90%.

3.7 Results
By constructing the timelines discussed above for every condition and by averaging the proportions of looks towards the target over the critical time windows, we were able to carefully examine data across conditions. We found that the evidence of apparent understanding varied somewhat predictably across conditions and ages as to where in time it was strongest, but that the 1 second window following the offset of the second target utterance was representative of the any general effect found in each condition. Thus while the data presented below all come from only this critical window, the effects compared are representative of the entire conditions. A 1 second window was chosen because this length of time gives the child enough time to reveal their understanding of the question but not so much time that they begin looking in a less systematic way. In no condition did we find effects of sex of infant, individual verbs or order of presentation, so these factors are not included in the analyses we report here.

### 3.7.1 15-month-olds

The mean looking time to the subject in both the Subject and Object condition after the second question across all trials is shown in Figure 1. Dividing this data into two blocks, the first three trials and last three trials, reveals an interesting pattern (Figure 1).
In order to quantify the factors determining looking time in this experiment more precisely, we built a series of candidate linear mixed effects models. These models, corresponding to alternate hypotheses about the effect of the block considered (all vs. first block vs. 2\textsuperscript{nd} block) and extraction type (Subject vs. Object) to infants’ looking times were fit in R (R development core team, 2008) with the \textit{lmer} function from the \textit{lme4} library (Bates, 2005; Bates and Sarkar, 2007) using maximum likelihood. The models were then compared using the \textit{anova} function in order to determine whether adding factors explained significant additional variance (Baayen 2007). The set of models that we compared are given in Table 2. Model 1 considers only the effect of block. Model 2 adds a term for the effect of the extraction type, independent of block. Model 3 includes both of these effects and an interaction term.
Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>Block</td>
<td>Subject</td>
</tr>
<tr>
<td>m2</td>
<td>Block + Extraction</td>
<td>Subject</td>
</tr>
<tr>
<td>m3</td>
<td>Block + Extraction + Block:Extraction</td>
<td>Subject</td>
</tr>
</tbody>
</table>

The analysis of variance comparing these models indicates that m3 is more explanatory than m1 or m2 ($\chi^2 = 17.47$, $p < 0.0002$). Finally, looking into m3 more closely with a traditional analysis of variance, we found a significant interaction of Block and Extraction ($F(2,90) = 5.9374$, $p < 0.005$). Focusing on the results from the 2nd Block, planned comparisons showed that the proportion of looks towards the subject varied significantly across extraction type (one tailed: $t(30) = 2.28$, $p < 0.2$), and mean proportion of looks towards the target in both subject questions and object questions differed marginally significantly from chance (one tailed: $t(15) = 1.40$, $p < 0.10$ and $t(15) = -1.98$, $p < 0.05$ respectively).

Interpreting the results from the first block (Figure 1) is less straightforward, as infants tend to look significantly more at the non-target character in both conditions. This appears to be some sort of baseline effect, however, as it is strong from the onset of the test sequence, even before the target question has been uttered (Figure 2). Additionally, we can see that despite this baseline effect, looks begin to pattern like the 2nd block (as predicted) by the last utterance of the target question.
3.7.3 20-month-olds: *Wh*-questions

With the 20-month-olds’ data we once again analyzed the mean looking time towards the subject in each condition and each block of trials separately (Figure 3).

Figure 2: Baseline ‘backwards’ behavior in 1st Block

Figure 3: Mean looking time to subject across blocks of trials
As with the 15-month-olds’ data, we wanted to quantify the factors determining looking time in this experiment more precisely and built the same series of candidate linear mixed effects models (Table 3).

**Table 3**

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>Block</td>
<td>Subject</td>
</tr>
<tr>
<td>m2</td>
<td>Block + Extraction</td>
<td>Subject</td>
</tr>
<tr>
<td>m3</td>
<td>Block + Extraction + Block:Extraction</td>
<td>Subject</td>
</tr>
</tbody>
</table>

The analysis of variance comparing these models indicates that m3 is more explanatory than m1 or m2 ($\chi^2 = 13.2682, p < 0.0002$). Finally, looking into m3 more closely with a traditional analysis of variance, we found a marginally significant interaction of Block and Extraction ($F(2,90) = 2.8526, p = 0.063$).

Planned comparisons in the 20-month-old data showed no effect of extraction type across all trials (one tailed: $t(30) = 0.62; p > 0.25$), and the mean proportion of looks towards the target did not differ from chance in either condition (Figure 3). As with the 15-month-olds’ data, we then analyzed the 1st block and the 2nd block as individual groups. We found no effect in the 1st block (one tailed: $t(30) = -0.84$, Figure 3), but we did find a significant effect of condition in the 2nd block (one tailed: $t(30) = 2.40; p < 0.02$), and here the mean proportion of looks towards the target differed significantly from chance in both conditions (one tailed: $t(15) = 1.86; p < 0.05$ and $t(15) = -1.59, p < 0.07$ respectively).

**3.7.3 Discussion of Results**
Based on the results presented above, it looks as though 15-month-olds behave as though they understand both subject and object *wh*-questions. This suggests that the concerns cited with the methodology in the Seidl et al paper were responsible for the subject-object asymmetry in 15-month-olds’ comprehension in that work. The fact that our effect was evident only in the last three trials strengthens the argument that the small number of trials in the Seidl et al study did not give 15-month-olds the opportunity to fully exhibit their comprehension abilities. These results suggest that 15-month-olds have the knowledge necessary to comprehend *wh*-questions, but that they are only able to properly deploy this knowledge under optimal conditions. As predicted, 20-month-olds behaved as though they understand both subject and object *wh*-questions; Their systems of knowledge and deployment are more solidly aligned with one another.

It is important to keep in mind that the fact that we were able to make the subject-object asymmetry disappear in 15-month-olds does not argue against the existence of such an asymmetry. The fact that it was object questions and not subject questions that broke down under suboptimal conditions reveals that 15-month-olds’ comprehension abilities for object questions are still more fragile than their abilities with subject questions. We explore the source of this fragility in experiment 2.

4 Experiment 2: Relative Clauses

4.1 Motivation
Although issues with the methodology appeared to underlie 15-month-olds’ asymmetrical performance on subject and object *wh*-questions in Seidl et al, the question remains as to why the asymmetry went in the direction that it did in previous work. That is, why, when experimental conditions were not ideal, were subject questions easier to comprehend than object questions? To probe this question we examined the comprehension of an arguably more difficult filler-gap dependency, the relative clause, using the same methodology as in experiment 1, which did not elicit an asymmetry in *wh*-questions. Thus target utterances were patterned after those in (20):

(20) Subject Relative Clause: Show me the dog that bumped the cat

Object Relative Clause: Show me the dog that the cat bumped

### 4.2 Predictions

Several predictions arise when testing the comprehension of relative clauses. First, if the asymmetry in the Seidl et al study could be resurrected with the more complicated Relative Clause structure, this structure would also be useful for disentangling the two linguistic hypotheses in 15-month-olds. That is, if the subject-object asymmetry stemmed from the longer structural distance between the filler and the gap in the object questions, it should maintain in relative clauses, where the gap is far from the filler. Alternatively, if the presence of *do*-support in the object questions was at the root of the asymmetry, it should disappear in relative clauses. Of course, it could be the case either that relative clauses are so much more difficult than *wh*-questions that no evidence of
their comprehension can be observed, or that relative clauses are not significantly harder than *wh*-questions, in which case no asymmetry might be expected.

### 4.3 Participants

32 15-month-olds (17 males) with a mean age of 14;27 (range: 14;04 to 15;17) and 32 20-month-olds (18 males) with a mean age of 20;03 (range: 19;10 to 20;29) were included in the final sample. Participants were recruited from the greater College Park, MD area and were acquiring English as native language. Parents completed the MacArthur Short Form Vocabulary Checklist. 15-month-olds' mean production CDI-vocabulary was (24.7) (range: 0 to 190), and 20-month-olds’ mean production CDI-vocabulary was (107) (range: 9 to 381). We analyzed the data of infants that completed at least 4 out of 6 test trials, and the trials where the infant was looking at least 20% of the time. Ten additional infants were tested but ultimately excluded from the analysis due to fussiness or inattention.

### 4.4 Materials and Procedure

The materials and procedure for Experiment 2 were identical to those of Experiment 1. The test phase of the stimuli was 2 seconds longer, after pilot data suggested that the amount of time given to respond to each question was not enough when the target utterance was lengthened.
4.5 Results

The results of Experiment 2 were analyzed in exactly the same way as those of Experiment 1. The data compared below comes from the 1 second window following the offset of the second target utterance, the same window as the data discussed from Experiment 1.

4.5.1 15-month-olds

As in Experiment 1, we analyzed the mean looking time towards the subject in each condition and each block of trials separately (Figure 4).

![Figure 4: Mean looking time to subject in after second target utterance](image)

As in Experiment 1, we wanted to quantify the factors determining looking time in this experiment more precisely and built the same series of candidate linear mixed effects models (Table 4).
Table 4

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>Block</td>
<td>Subject</td>
</tr>
<tr>
<td>m2</td>
<td>Block + Extraction</td>
<td>Subject</td>
</tr>
<tr>
<td>m3</td>
<td>Block + Extraction + Block:Extraction</td>
<td>Subject</td>
</tr>
</tbody>
</table>

The analysis of variance comparing these models indicates that m2 is more explanatory than m1 ($\chi^2 = 9.6277, p < .0002$), and m3 is more explanatory than m2 ($\chi^2 = 7.3521, p < .03$). Finally, looking into m3 more closely with a traditional analysis of variance, we found a marginally significant effect of Block (F(2,90), p<.09), a significant effect of Extraction (F(1,90), p < .0001) and a marginally significant interaction of Block and Extraction (F(2,90), p = .10).

Planned comparisons showed an effect of extraction type across all trials (one tailed: t(30) = 2.78, p < 0.005), and in both extraction types the mean proportion of looks toward the target differed significantly from chance (one tailed: t(15) = 2.15, p < 0.03 and t(15) = -1.76, p < 0.05) (Figure 4). In the 1st block, the effect of extraction type disappeared and only the subject condition differed significantly from chance (one tailed: t(15) = 1.75, p < 0.06) (Figure 4). In the 2nd block, however, the effect of extraction type was apparent (one tailed: t(30) = 3.85, p < 0.001) and the mean proportion of looks toward the target differed significantly from chance in both conditions (one tailed: t(15) = 1.75, p < 0.06 and t(15) = -3.56, p < 0.002) (Figure 4).
Because it appears that the 15-month-olds can comprehend both subject and object relative clauses, these results cannot tell us which of the linguistic hypotheses, *do*-support of structural distance, lay behind the asymmetry in the Seidl et al paper. It is either the case that *do*-support was the problem, or that relative clauses are not difficult enough to elicit the asymmetry. The 15-month-olds’ success becomes interesting, however, when we see that it does not parallel 20-month-olds’ behavior on the same task.

### 4.5.2 20-month-olds

As with the 15-month-olds’ data, we analyzed the mean looking time towards the subject in each condition and each block of trials separately (Figure 5).

![Figure 5: Mean looking time towards subject following the 2nd question](image)

As in Experiment 1, we wanted to quantify the factors determining looking time in this experiment more precisely and built the same series of candidate linear mixed effects
models (Table 5).

Table 5

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>Block</td>
<td>Subject</td>
</tr>
<tr>
<td>m2</td>
<td>Block + Extraction</td>
<td>Subject</td>
</tr>
<tr>
<td>m3</td>
<td>Block + Extraction + Block:Extraction</td>
<td>Subject</td>
</tr>
</tbody>
</table>

The analysis of variance comparing these models indicates that neither m1 nor m2 are better at explaining the variance than m1 ($\chi^2 = 0.8030$, $p = 0.3702$ and $\chi^2 = 1.2337$, $p = 0.5396$ respectively. Planned comparisons showed no effect of condition across all trials and the mean proportion of looks toward the target did not differ significantly from chance in either condition (Figure 5). We split up the trials into the 1st and 2nd blocks and analyzed again, and in neither block did any condition differ significantly from chance (Figure 5).

4.6 Discussion of Results

This U-shaped pattern of results, 15-month-olds seem to successfully interpret both subject and object relative clauses, but 20-month-olds appear unable to comprehend either type of relative clause, is unexpected for two reasons. First, the grammatical parallels between wh-questions and relatives leads us to expect the processing of wh-questions and relative clauses to rely on the same mechanisms. Consequently, if an age group is able to process one of these constructions, we would expect them to be able to process the other. Second, it is unexpected that older infants, who are presumably more grammatically advanced than younger infants, would not be able to understand something
that younger infants and adults can. This U-shaped pattern suggests a disjunct between the development of knowledge and the necessary deployment systems for this knowledge between 15 and 20 months of age.

5 General Discussion

Looking at our results with respect to the knowledge-deployment distinction introduced above, we can begin to make sense of our finding of U-shaped development. In each age group, we ultimately want a characterization of both the way that children represent the filler-gap dependencies, which we will abbreviate as Kn, where K stands for “knowledge” and n=age, and the parsing mechanisms that enable them to deploy their knowledge, which we will represent as Dn, where D stands for “deployment” and n=age. Based on 15-month-olds' roughly parallel performance with relative clause and wh-questions, we can assume that whatever knowledge state (K15) they are in regarding filler-gap dependencies, their deployment system (D15) is appropriately matched. That is, the deployment system D15 accurately delivers information compatible with the knowledge state K15 for both subject and object dependencies, independent of the wh/relative distinction. However, due the differences in performance we see in the 20-month-olds we know that there must be some disjunct between K15 and K20, D15 and D20 or both. Below we will characterize the knowledge states and deployment systems at 15 and 20 months in order to identify a developmental trajectory in line with the empirical findings.
5.1 U-shaped development

One assumption about language acquisition (maybe development in general) will be critical to the following discussion. Upon reaching a mature or adultlike knowledge state, children do not regress to less complete knowledge states and later recover from these losses to be truly adultlike. While there are cases that may look like a temporary loss of knowledge, known as U-shaped development, such cases are best characterized by a growth in knowledge accompanied by a delay in the acquisition of deployment systems compatible with that knowledge (Pinker 1984). Initial success in the U-shaped pattern can be characterized as incomplete knowledge yielding erroneous success. The dip in performance can be characterized as a growth of knowledge accompanied by a deployment system that lags behind. Finally, the third piece of the U, where children regain their success, reflects the integration of the knowledge and deployment systems.

To make this characterization more concrete, let us consider one of the most famous cases of U-shaped development: the acquisition of English past tense morphology (Pinker 1984). In this case, very young children produce frequent irregular past tense forms fairly accurately. This could be characterized as their knowledge state being one that has no rules for forming past tense, but simply a memorized list of past tense forms. When their knowledge states changes into one that includes the rule of *+ed* for past tense morphology, their deployment system also shifts, so that instead of simply retrieving stored forms, it applies to past tense rule to every verb retrieved. Their knowledge of the stored form for irregulars hasn’t changed, but their ability to deploy both the stored form and the rule has not yet reached the adult state (in which stored forms are accessed faster
than those produced by rule). The external effect of this shift is that children who appeared mature in producing correct irregular past tense forms now make many non-adult-like errors. But in fact, their knowledge system has not regressed. On the contrary, it has progressed, but due to the deployment system’s need to adapt for the new knowledge state, this looks like a regression in knowledge. Eventually, the deployment state is able to access both the stored irregular forms and the past tense rule, and the child produces past tense forms like an adult would. We now propose that the development of the parsing mechanisms necessary for filler-gap dependencies can be looked at in a parallel fashion.

5.2 Deriving the U-shaped trajectory

To derive the U-shaped trajectory in the development of the parsing mechanisms necessary for filler-gap dependencies we have a set logical of possibilities to choose from. It is a matter of logical necessity that one of the scenarios in (21) must be true:

(21) (a) $K_{20} = K_{15}$, $D_{20} = D_{15}$
    (b) $K_{20} \neq K_{15}$, $D_{20} = D_{15}$
    (c) $K_{20} = K_{15}$, $D_{20} \neq D_{15}$
    (d) $K_{20} \neq K_{15}$, $D_{20} \neq D_{15}$

As mentioned above, any time a knowledge state ($K_n$) is not equal to its predecessor ($K_{n-1}$) we can assume that the later state is a more advanced one, subsuming the knowledge
from the earlier state. We can rule out scenario (a) on the grounds that if $K_{20}$ were equal to $K_{15}$, and $D_{20}$ were equal to $D_{15}$ then we would not expect to see any difference in behavior. Since we do see a difference, we know that some change must take place. At this point we can also rule out scenario (b) as it is not clear why, if knowledge is increasing between $K_{15}$ and $K_{20}$, and if the deployment system remains the same, comprehension would decrease. These deductions leave us with two scenarios, (c) and (d). The difference between these depends on whether $K_{15}$ is equal to $K_{20}$ or not.

In order to determine whether there is a change in the relevant knowledge state between 15 and 20 months it is first necessary to determine what these knowledge states could be. Then we will need to describe the deployment systems that would go along with these knowledge states and see which derivation yields the observed U-shaped development. We will begin by characterizing $K_{20}$.

5.2.1 Characterizing $K_{20}$

There are two hypotheses we can make about the nature of $K_{20}$. It is either an adultlike knowledge state or a state intermediate between $K_{15}$ and an adultlike state. In the case that their knowledge is adultlike, children’s grammars would treat the formation of $wh$-questions and relative clauses as exhibiting parallel structures. If this were the case, difficulty in comprehending relative clauses would come from problems with the deployment of this knowledge. If 20-month-olds are at some intermediate knowledge state, it would be one where $wh$-questions are not represented in the same way as relative
clauses. In this case they would presumably have a deployment system that works for
*wh*-questions but not relatives.

As a simplifying assumption we will tentatively characterize 20-month-olds’
knowledge as adultlike. Because evidence for the child’s knowledge state is indirect, and
we can only know as much as their deployment system can show us, it is a simpler
developmental hypothesis to believe that children’s representations are continuous with
adults’ unless we find clear evidence to the contrary. Also, at no later stage in acquisition
do we have evidence that *wh*-questions and relative clauses are treated differently, nor do
we have any hypothesis as to how children would reunite them as instances of one
phenomenon after splitting them up. Finally, while we will not present it here, recent
results from our lab show that 20-month-olds can comprehend relatives clauses when
processing demands are reduced, suggesting that they do have the appropriate knowledge
for relativization but have difficulty deploying this knowledge. Thus we tentatively
characterize $K_{20}$ as an adultlike knowledge state, including the grammatical constraints on
the formation of both types of filler-gap dependencies.

5.2.2 Is $K_{15}$ equal to $K_{20}$?

If $K_{15}$ is equal to $K_{20}$ it would mean that as early as 15-months children have knowledge
of syntactic movement and resolution of filler-gap dependencies. With similar knowledge
states, the difference we see in behavior between 15- and 20-months would have to be
derived from differing deployment systems. That is, we are dealing with scenario (c),
where $K_{15} = K_{20}$ but $D_{15} \neq D_{20}$. As mentioned above in Section 2.1, it is widely believed
that adults employ an active filling strategy to resolve filler-gap dependencies. Because of this we will take $D_{20}$ to be an active filling strategy, resolving dependencies by positing gaps once a filler has been encountered. The only option for $D_{15}$ that could correspond to the grammatical knowledge posited, would be a gap-driven parser. That is, unlike adults or 20-month-olds, these children do not begin to form a filler-gap representation until they encounter a gap. Upon finding one (in either *wh*-questions or relatives), they would search back through the sentence to find the filler.

Now that the relevant knowledge states and deployment systems have been characterized, we can see if they derive the U-shaped development. While not adultlike, it is apparent that gap-driven parsing could lie behind 15-month-olds’ success with both *wh*-questions and relative clauses. We would expect a 20-month-old with a developed active filling strategy to succeed with both dependency types as well and this does not predict the observed pattern of behavior. However, we might expect that a child developing an active filling strategy to succeed when processing demands are lower, and fail when they are higher. Taking into consideration the superficial variation between *wh*-questions and relative clauses we can see several differences that could make relative clauses more difficult to parse with an active filling strategy. In a *wh*-question, (a) the filler is sentence initial, (b) the filler is morphologically marked by virtue of being a *wh*-word and (c) *wh*-questions exhibit a marked question intonation that is lacking in relatives. All of these features are signals to the listener that could signal that a filler-gap dependency needs to be resolved, but are not available in the relative clauses that we used.
While this explanation could explain what 15- and 20-month-olds’ deployment systems are, it is as yet unclear why 15-month-olds would shift away from gap-driven parsing if it is working for them. One idea is that as time goes by, the child would take note of a statistical correlation between the appearance of *wh*-words (or other fillers) and the subsequent identification of a gap. Between 15 and 20 months the child must adapt the deployment system to one that posits gaps after finding *wh*-words and can relate the two (D20).

This could explain how the child shifts from being a gap-driven parser to an active filler, but it does not predict that there will be a point in time when the child cannot resolve certain filler-gap dependencies. That is, even once the correlation between filler and gap is noted and used predictively, there is nothing incompatible in this strategy with falling back on gap driven parsing when active filling fails. In the case where the filler is less salient, such as in our relative clause condition, the child should still eventually come upon the gap and be able to look back to find the filler. While this might predict a difference in the time course of resolution of different construction types, it doesn’t predict the behavior we’ve found, where relatives appear to be unresolvable.

The developmental trajectory predicted by a constant knowledge state and a transition from gap driven parsing to active filling is shown in Figure 5. Because this trajectory doesn’t match up with the behavioral results, we maintain our conclusion that $K_{15} \neq K_{20}$ but reject the hypothesis that $D_{15}$ is gap-driven parsing.
5.2.2 $K_{15}$ is not equal to $K_{20}$

Returning to our original four possible scenarios (22), we have now ruled out (a), (b) and (c). We are left to see whether (d) results in the derivation of the U-shaped trajectory.

(22) 
(a) $K_{20} = K_{15}$, $D_{20} = D_{15}$

(b) $K_{20} \neq K_{15}$, $D_{20} = D_{15}$

(c) $K_{20} = K_{15}$, $D_{20} \neq D_{15}$

(d) $K_{20} \neq K_{15}$, $D_{20} \neq D_{15}$
As above, we will assume that $K_{20}$ is an adultlike knowledge state. We can account for $D_{20}$ in the same way as above as well, being an active filling strategy that struggles under higher processing demands. What is left is to characterize $K_{15}$ and $D_{15}$.

If $K_{15}$ must be different than $K_{20}$, then it must not involve knowledge of filler-gap dependencies or syntactic movement. What 15-month-olds do know about, however, are verb meanings (Golinkoff et al 1995). Knowing the meaning of a verb implies knowledge of the thematic roles associated with a given verb. This knowledge in turn implies knowing that transitive verbs require two participants. $D_{15}$ then, would be a heuristic parsing strategy that relies on knowledge of argument structure, and relatedly, event structure, instead of syntactic dependencies. The heuristic depends on the identification of a verb missing a noun phrase needed to fill a required thematic role. The child would recognize a gap in the argument structure by noticing a substring in which an expected syntactic argument fails to occur (e.g., the cat bumped __ in a filler-gap dependency involving an object). Having identified a verb that is missing a required argument, the heuristic parser would then search the discourse context for a referent that could fill out this thematic structure. It is important to note that if 15-month-olds are relying on this heuristic they are crucially not making the link between the filler and the gap, and do not even need to parse or interpret the filler to arrive at the correct interpretation.

Once again, we must determine why children would ever abandon this strategy if it works as well as it appears to. We have already posited that by 20-months child have the appropriate grammatical structure and constraints to be able to interpret syntactic movement. In particular, they have learned about the relation between movement and
subcategorization, realizing that a verb can sometimes find its arguments in displaced positions in the clause. It follows that once this system is in place, extragrammatical heuristics like the one proposed above wouldn’t be available to parse these sentences because of the grammatical constraint requiring that subcategorized arguments must be syntactically realized. At this point, infants will need to develop a new system to deploy their updated knowledge of filler-gap dependencies: active filling ($D_{20}$).

This account predicts that during the transition period, once the heuristic ($D_{15}$) has been rendered incompatible by the updated knowledge ($K_{20}$), the child will begin to use an active filling strategy ($D_{20}$). As discussed above, a developing active filler is likely to find certain fillers more difficult to identify, such as those in our relative clause condition, and poorer performance is predicted. Unlike in the transition from gap-driven parsing to active filling, once $D_{20}$ comes online, $D_{15}$ is no longer compatible with the $K_{20}$ and cannot be used as a backup mechanism when active filling fails. This trajectory (schematized in Figure 6) lines up very well with our empirical results.
Figure 6: Schematic of the development of knowledge and deployment from a heuristic system to an adultlike one.

5.3 Predictions for Future Work

The account given above makes several predictions regarding the comprehension of different types of filler-gap dependencies by both 15- and 20-month-olds. First, since we are proposing that 15-month-olds are not using the filler when they comprehend sentences with filler-gap dependencies, they should not make distinctions dependent on information in the filler. For example, if they were presented with a situation where a cat bumped a boy, the cat bumped the a truck and then a girl bumped the cat, and then asked Who did the cat bump?, they should be able to narrow down the choices to the two possible object, the boy and the truck, but should not differentiate between them, despite the fact that an adult using the filler, and by hypothesis a 20-month-old, would use the animacy restriction on who to choose the boy.

Second, since 15-month-olds are only using knowledge of thematic roles, not the structure of the dependency, to resolve the missing argument, they should not be sensitive to illicit extractions that adults are, and 20-month-olds should be. For example, given the situation outlined above, for an adult the utterance What did the cat bump and the boy? would be ungrammatical as a violation of the coordinate structure constraint. While an interpretation might ultimately be reached, it might not follow the time course of licit question answering. If a 15-month-old were only filling in thematic structure with an
appropriate referent, they might be able to choose the appropriate referent in a manner similar to answering a licit question.

Finally, regarding the 20-month-olds’ failure with relative clauses, we would predict that having a more salient filler, i.e. a wh-relative such as Show me the dog who bumped the cat they would have less trouble identifying the presence of a filler and subsequently resolving the dependency. We are currently testing all of these predictions in our lab (Gagliardi & Lidz 2010).

5.4 Theoretical Implications

If the account given above proves to be an accurate characterization of the development of filler-gap dependencies, it also provides the beginnings of an argument against parsing models which do not use details of grammatical representation to build sentence interpretations, as in the models of ‘good-enough’ parsing illustrated by Ferreira et al (2002) or Townsend & Bever (2001). These views suggest that the parser computes interpretations of sentences using heuristics that yield interpretations similar to those that would be derived by a system that uses grammatical detail in real time. This kind of model is similar to what we posit for 15-month-olds, but doesn’t account for why 20-month-olds would stop using this strategy, as it should still work. Consequently, if the asymmetry at 20-months in the comprehension of wh-questions and relative clauses derives from the combination of an adultlike grammar and an inefficient parser, it looks as though the parser does its best to implement the grammar and does not settle on a good-enough parse. That is, while the good-enough view could account for 15-month-
olds’ behavior, it doesn’t appear to ultimately characterize the interaction between the grammar and the parser in development.

6 Conclusion

In this paper, we have identified a case of U-shaped development in the domain of filler-gap dependencies. Whereas 15-month-old children seem to correctly interpret both subject and object wh-questions and subject and object relative clauses, 20-month-olds seem to have lost the ability to correctly interpret relative clauses. We have proposed that this developmental pattern can be explained in a framework that identifies independent contributions of (a) grammatical knowledge, (b) the information processing mechanisms that deploy that knowledge, and (c) the alignment of those mechanisms during language development. We have argued that in the case of filler-gap dependencies, both knowledge and deployment vary across development. We have proposed that 15-month-olds have impoverished grammatical representations for these dependencies and that their deployment systems are appropriate for those representations. Twenty-month-olds, on the other hand, have accurate adult-like knowledge but have yet to become effective at deploying that knowledge in real-time.
References


Appendix A: Descriptions of stimuli

Verbs (participants)

Bump (white dog, cat, brown dog)
Kiss (brown monkey, goose, black monkey)
Hug (frog with hat, bear, frog with scarf)
Wash (brown monkey, elephant, black monkey)
Tickle (white mouse, bee, gray mouse)
Feed (frog with hat, elephant, frog with scarf)

Experiment 1: WH-Questions

Subject Condition / Object Condition

Which dog bumped the cat? / Which dog did the cat bump?
Which monkey kissed the goose? / Which monkey did the goose kiss?
Which frog hugged the bear? / Which frog did the bear hug?
Which monkey washed the elephant? / Which monkey did the elephant wash?
Which mouse tickled the bee? / Which mouse did the bee tickle?
Which frog fed the elephant? / Which frog did the elephant feed?

Experiment 2: Relative Clauses

Subject Condition / Object Condition

Show me the dog that bumped the cat / Show me the dog that the cat bumped
Show me the monkey that kissed the goose / Show me the monkey that the goose kissed
Show me the frog that hugged the bear / Show me the frog that the bear hugged
Show me the monkey that washed the elephant / Show me the monkey that the elephant washed

Show me the mouse that tickled the bee / Show me the mouse that the bee tickled

Show me the frog that fed the elephant/ Show me the frog the elephant fed