Discontinuous Development in the Acquisition of Filler-Gap Dependencies: Evidence from 15- and 20-Month-Olds

Annie Gagliardi, Tara M. Mease & Jeffrey Lidz

To cite this article: Annie Gagliardi, Tara M. Mease & Jeffrey Lidz (2016): Discontinuous Development in the Acquisition of Filler-Gap Dependencies: Evidence from 15- and 20-Month-Olds, Language Acquisition

To link to this article: http://dx.doi.org/10.1080/10489223.2015.1115048

Published online: 13 Apr 2016.

Submit your article to this journal

View related articles

View Crossmark data
Discontinuous Development in the Acquisition of Filler-Gap Dependencies: Evidence from 15- and 20-Month-Olds

Annie Gagliardi a, Tara M. Mease b, and Jeffrey Lidz b

aHarvard University; bUniversity of Maryland

ABSTRACT
This article investigates infant comprehension of filler-gap dependencies. Three 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, while Experiment 3 shows that when parsing demands are reduced, older children can comprehend them as well. We argue that this discontinuous pattern follows from an offset in the development of grammatical knowledge and the deployment mechanisms for using that knowledge in real time. Fifteen-month-olds, we argue, lack the grammatical representation of filler-gap dependencies but are able to achieve correct performance in the task by using argument structure information. Twenty-month-olds do represent filler-gap dependencies but are inefficient in deploying those representations in real time.

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 (Miller & Chomsky 1963; Chomsky 1965; Bloom 1990; Crain & Thornton 1998; Trueswell et al. 1999). Studies of children’s on-line comprehension abilities typically focus strictly on deployment processes, holding grammatical knowledge fixed (Trueswell et al. 1999; Swingley, Pinto & Fernald 1999; Arnold et al. 2004; Omaki et al. 2014), whereas researchers interested in the character of children’s grammatical representations typically treat performance factors as noise variables whose effects need to be factored out for the purposes of assessing knowledge (Hamburger & Crain 1984; Bloom 1990; Musolino & Lidz 2006; Vial, Lidz & Musolino 2010). The current article weds these perspectives together by examining the development of filler-gap dependencies in children in the second year of life. Maintaining the view that real-time understanding requires deploying the child’s current linguistic knowledge, we explore the possibility that a child with immature linguistic knowledge may nonetheless display appropriate interpretive behavior because of the way that partial knowledge is deployed. Moreover, we show that certain errors in comprehension derive from aspects of the online performance systems. These effects yield a discontinuity in development, with 15-month-olds succeeding with all filler-gap dependencies and 20-month-olds succeeding with wh-questions but not relative clauses. These effects emphasize the importance of considering the contributions of knowledge and deployment simultaneously when assessing grammatical knowledge and accounting for its growth in a learner.

In this article we investigate the relation between the acquisition of grammatical knowledge and the accompanying deployment systems. Using filler-gap dependencies as a probe, we explore children’s understanding across two ages in order to hone in on the source of their successes and
failures. We focus on children at 15 and 20 months of age because these children are at the very beginning stages of acquiring the syntax of their native language and so provide a likely testing ground for examining the relation between early grammatical knowledge and deployment systems. Specifically, we compare wh-questions and relative clauses because these structures rely on the same core grammatical representations but diverge in the deployment processes they engage because of differences in surface form.

Due to both their unbounded nature and uniquely linguistic character, filler-gap dependencies provide an ideal place to examine the relation between grammatical knowledge and deployment. In our investigation of the acquisition of these two types of dependencies, we observe a case of discontinuous development that can be explained in terms of growth of grammatical knowledge with a delay in the real-time deployment mechanisms. The basic empirical finding of the article is that 15-month-olds successfully interpret wh-questions and relative clauses, but 20-month-olds show success only with wh-questions. To explain this pattern, we argue that initial success with both constructions is based on partial grammatical knowledge, but the asymmetry between wh-questions and relatives at 20 months reflects acquisition of the relevant grammatical knowledge, paired with difficulty deploying that knowledge in the case of relative clauses. The latter conclusion is supported by additional experimentation showing an asymmetry between that-relatives and wh-relatives in 20-month-olds.

2. Background: Filler-gap dependencies

Filler-gap dependencies relate an element in a nonthematic position (henceforth the “filler,” shown in italics) to its canonical thematic position in the sentence (henceforth the “gap,” marked by ____). 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 ____?

Wh-questions (1) are only one type of filler-gap dependency. Another structure, the relative clause (3), is also a filler-gap dependency.

(3) Show me the dog that the cat bumped ____

There are several differences in the surface properties of wh-questions (1) and relatives (3): the presence of a wh-word in (1) and its absence in (3); 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; the fact that sentences containing relative clauses are necessarily multiclausal while those containing wh-questions need not be; and, when uttered aloud, the different prosodic contours associated with the two structures. 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 (4–6) (Chomsky 1977).

(4) 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 ____?
(5) a. *Which dog did the monkey think that ____ bumped the cat?
   b. *Show me the dog that the monkey thought that ____ bumped the cat
(6) 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

---

1We restrict our attention in this article to dependencies involving arguments.
gap. Following is an overview of the processes 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, Phillips & Weinberg 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 cross-linguistic 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 compatible with the semantic information found in the filler (16) (Stowe 1986; Traxler, Morris & Seely 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

In order for active filling to be successful, fillers have to be reliably identified and maintained (or reaccessed) in memory until the gap site is encountered. If a comprehender were to miss a filler, they wouldn’t know to posit a gap and would subsequently fail with the relevant dependency resolution.

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; Villiers, Roeper & Vainikka 1990; Stromswold 1995; Thornton 1995). Children have been found to produce these dependency types as young as 1;08 (Stromswold 1995). Of course, if our goal is to look at the very beginning of the acquisition of these structures (the point in time when children can first recognize that a sentence involves such a dependency and deploy knowledge of this dependency to successfully comprehend it), we must look at comprehension abilities before children are producing these sentences. While comprehension of both wh-questions and relative clauses has been studied extensively (Tavakolian 1981; Goodluck and Tavakolian 1982; Hamburger & Crain 1982; Roeper & de Villiers 1994; Corrêa 1995; Mckee, McDaniel & Snedeker 1998; McDaniel, Mckee & Garrett 2010; Kidd 2011; Villiers & Roeper 1995; Friedmann, Belletti & Rizzi 2009; Goodluck 2010; among many more) most of this work looks at comprehension abilities in children much older than 1;08. Only one study that we know of examines children young enough that we might learn something about their comprehension abilities at the very earliest stages of syntactic acquisition.

In important foundational work, Seidl, Hollich, and Jucszyk (2003) used the intermodal preferential looking procedure to examine comprehension of wh-questions by 13-, 15-, and 20-month-olds. This age range was tested because it looks as though the earliest production of wh-questions is at about 20 months (Stromswold 1995), and as mentioned earlier, we would expect comprehension to precede production. Furthermore, we might expect that asymmetries in comprehension of subject and object extraction reveal something about how these structures are learned. Each infant was
tested on the comprehension of two subject questions, two object questions, and one where question. They found that 20-month-olds appeared to understand all three question types, 15-month-olds appeared to understand only subject and where questions, and 13-month-olds did not appear to understand any question type.

These authors 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. The difficulties displayed by 15-month-olds in Seidl, Hollich, & Jusczyk (2003) thus indicate that at a much earlier age than has been tested before children do have some knowledge of filler-gap dependencies, but their abilities at this stage are fragile. This fragility could be grammatical in nature or derived from aspects of their deployment mechanisms. The lack of these difficulties at 20 months points toward an explanation of the asymmetry that is different from that found in adults and other children. That is, rather than believing that some aspect of grammatical knowledge or processing causes an asymmetry at 15 months, disappears at 20, but then resurfaces in preschool, we will investigate whether the asymmetry could have a different root at the two stages. Thus the developmental changes that take place between 15 and 20 months provide the central motivation for the current experiments, as we extend the findings of Seidl, Hollich, & Jusczyk and probe the knowledge of 15-month-olds more closely.

3. Experiment 1: wh-questions

3.1. Motivation

Our first goal is to determine whether a lack of knowledge or an inability to deploy knowledge lies behind the difference between 15-month-olds’ and 20-month-olds’ performance in the Seidl, Hollich & Jusczyk (2003) study. To do so, we first need to take a closer look at their study. While Seidl, Hollich & Jusczyk cited the longer structural distance and do support as the two factors that could have made object extraction too difficult for 15-month-olds, in doing so they assume that 15-month-olds entertain adultlike representations of wh-questions and that any problems they exhibit would be due to problems accessing these representations. That is, they align the subject-object asymmetry found in 15-month-olds with the performance asymmetries found in much older children and adults. What is curious about this account is that this asymmetry would show up at 15, but not 20, months. While it is possible that some common factor is affecting 15-month-olds’ and adults’ performance, this account leaves unexplained why the subject-object asymmetry would disappear in 20-month-olds, only to return at age 3 or 4. It strikes us, then, that 15-month-olds might not possess all the knowledge that adults do and that their difficulty with object wh-questions stems from some other source.3

Indeed, it seems somewhat unlikely that infants as young as 15 months would possess fully adultlike knowledge of filler-gap dependencies and be able to deploy this knowledge online using the same parsing mechanisms that give rise to the subject-object asymmetry in adults. Prior to 20 months, infants have not yet fully mastered the link between a word’s grammatical category and its meaning (Waxman & Booth 2003), show variable performance on tests of argument structure (Yuan, Fisher & Snedeker 2012; Lidz, White & Baier 2013), and fail to correctly interpret number marking on NPs (Kouider et al. 2006). Thus, it is not unreasonable to suppose that they have yet to determine which sentences of English should be represented as involving a filler-gap dependency. Of course

3An anonymous reviewer raises the possibility that the mechanisms that give rise to the subject–object asymmetry are present throughout development but that the preferential looking paradigm is not sensitive enough to detect it in 20-month-olds. This suggestion strikes us as reasonable. On the other hand, this asymmetry shows up in choice tasks that are very similar to ours in children older than 20 months (e.g., Brandt et al. 2009). Given the basic similarity between those tasks and ours and the general sensitivity of preferential looking to interpretation in children ranging from 2 months (Baier et al. 2007) through at least 30 months (Lukyanenko et al. 2014), we think there is also reason to doubt this very plausible possibility.
this is not to say that 15-month-olds lack the capacity to represent filler-gap dependencies (our research does not bear on this question), just that they may not yet know which strings of English require those representational resources.

Before we can investigate the character of 15-month-olds’ knowledge, we first have to determine what caused the asymmetry between 15- and 20-month-olds in Seidl, Hollich, & Jusczyk’s (2003) task. Here we explore whether methodological difficulties could have contributed. In their task, 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 silently 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 is highly unrealistic, given that two-dimensional inanimate objects do not normally move spontaneously.

It is also pragmatically odd to ask questions about such scenes. Because only one event took place, the question was pragmatically infelicitous. Why ask a question like “what hit the keys?” when there were only two objects present and one interaction between them?

These issues that potentially interfere with infants’ attention and memory during the task could impact infants’ performance, whether they are parsing these sentences as adults might or taking some other route to an interpretation. That is, if infants are parsing grammatically, and one parse is more difficult in general than another (e.g., object gap resolution as compared to subject gap resolution), the more difficult parse to achieve might be affected. But similarly, if children use an alternative mechanism that is better practiced in interpreting sentences with missing subjects, it could be easier to deploy this heuristic when attention and memory are taxed. It appears reasonable to believe that subject gap resolution is in fact more common in the child’s experience due to a higher incidence of missing subjects in contexts like imperatives, omitted subjects and auxiliaries in yes/no questions (wanna go?), and a large number of predicative questions (e.g., what’s that?).

Thus in Experiment 1 we set out to probe whether factors in the design and materials employed by Seidl, Hollich, & Jusczyk (2003) could have masked the infants’ underlying abilities. In this way we will be able to better gauge infants’ early comprehension abilities and from there begin to probe the source of these abilities.

In order to investigate this possibility, we made several modifications to the basic design of the Seidl, Hollich, & Jusczyk (2003) study. Target utterances were wh-questions patterned after those in (19):

\[
\begin{align*}
\text{(19) Subject WH Question: Which dog bumped the cat?} \\
\text{Object WH Question: Which dog did the cat bump?}
\end{align*}
\]

To explore this possibility, we attempted to improve upon the factors we identified as potentially problematic earlier. 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 trial, enabling us to determine whether having too few trials can obscure 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 whom, 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, which will become important later.

\[\text{Note that this is not the same as the picture-matching methodology that has been used with some success with older children. In our task, children see an event, then see the two participants side by side, and then need to pick one, unlike picture-matching tasks where participants see two pictures depicting different events or objects and then need to choose one.}\]
While this design allows us to improve upon Seidl, Hollich, & Jusczyk’s (2003) methodology, the use of preferential looking to probe complex syntactic knowledge is in itself a methodological experiment. That is, we do not yet know the best way to test for comprehension of this type of knowledge (or deployment of this knowledge). Their design seemed promising yet left much room for improvement. Ours is an improved version, and only through using it will we discover its own shortcomings.

From the outset we know that our task shares some potential shortcomings with that of Seidl, Hollich, & Jusczyk (2003). For example, both tasks require the infants to encode and remember an event and invoke their memory of this event to answer the question. We do not know what role this memory load will play, especially compared to other cases where knowledge of syntax has been examined more directly, via ongoing scenes that correlate with the linguistic material (e.g., tests of verb learning via syntactic bootstrapping [e.g., Naigles 1996, among others] and binding constraints [Lukyanenko et al. 2013]). Additionally, we were initially unsure whether infants would be able to process and answer the question immediately, whether they would need to hear several iterations of the question, or whether their language processing would benefit from the removal of distracting visual information. Thus we included three repetitions of the target question in our test phase, the second of which was uttered while the screen was blank, to limit the information the infant needed to integrate at that moment in time. We expected that if infants could comprehend the linguistic stimuli, this would be evident in their looking behavior in a time window following one of these questions, but not all three. Thus, this method represents an initial step toward finding the ideal conditions to probe children’s syntactic knowledge using preferential looking techniques.

3.2. Predictions

The predictions for this first experiment are straightforward. Regarding 15-month-olds, if the issues concerning felicity and engagingness were responsible for the subject-object asymmetry in the Seidl, Hollich, & Jusczyk (2003) experiment, then these asymmetries should disappear when these concerns have been addressed. In addition, if the difficulties introduced by these trial properties are amplified by the use of too few trials, then we predict an effect of trial number, with 15-month-olds showing greater success in later trials than in early trials. On the other hand, if the subject-object asymmetry was not caused by methodological issues, then we expect it to be maintained in the current design. Under either possibility, 20-month-olds are predicted to correctly interpret both subject and object *wh*-questions. Finally, as mentioned earlier, if infants can understand anything, we expected them to show evidence of this comprehension following one, but not necessarily all, of the utterances of the target question.

It is important to remember, however, that this experiment will not necessarily tell us whether children’s correct interpretations derive from their having the same grammar and using the same parsing mechanisms as adults. Rather, having clarified their interpretative abilities, we will be in a better position to ask about the source of those abilities.

3.3. Participants

Thirty-two 15-month-olds (16 males) with a mean age of 15.00 (range: 14.14 to 15.18) and thirty-two 20-month-olds (16 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, Maryland, area and were acquiring English as native language. Parents completed the MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al. 1993). The 15-month-olds’ mean production CDI vocabulary was (19.2) (range: 0 to 60, out of a total possible 655), and the 20-month-olds’ mean production CDI vocabulary was (125) (range: 21 to 574, out of a total possible 655). We analyzed the data of infants who completed at least 4 out of 6 test trials (63/64 infants analyzed watched 6/6 test
trials), and the trials where the infant was looking at least 20% of the time (this excluded 6 trials). 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 digital video recordings of puppets performing the actions on one another. This footage was edited to create the series of events outlined in Table 1. All sequences were filmed against a white background and presented on a 51” plasma television screen. We don’t show stills here as they do not capture the dynamic nature of the videos, but a sample video of an entire trial can be found at (ling.umd.edu/labs/acquisition/stimuli/wh_s_bump.mp4).

#### 3.4.2. Auditory stimuli

The audio portion of the stimuli (as outlined in Table 1) was 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 parental permission. The infant and parent were then escorted into a soundproof room, where the infant was either seated on the parent’s lap or in a high chair, centered 6 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

<table>
<thead>
<tr>
<th>Table 1. Schematic of One Entire Trial.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length of</strong></td>
</tr>
<tr>
<td><strong>segment (s)</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>4.67</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>4.00</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>4.00</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>4.00</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>4.00</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>7.00</td>
</tr>
<tr>
<td>3.00</td>
</tr>
<tr>
<td>7.00</td>
</tr>
<tr>
<td>1.00</td>
</tr>
<tr>
<td>6.00</td>
</tr>
<tr>
<td>3.33</td>
</tr>
<tr>
<td>6.00</td>
</tr>
</tbody>
</table>

Audio segments marked by an asterisk (*) varied as a function of condition.
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). All of the six verbs chosen are words that at least 37% of 15-month-olds (average 56%) are expected to know based on comprehension data from the Lex2005 database (Dale & Fenson 1996) (see the appendix for complete descriptions). To focus infants’ attention before the beginning of each trial, a 4-second still image 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 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
At 20 seconds infants were introduced to each of the animals that would be involved in the action (4 seconds each, followed by a 1-second black screen break), and then shown a shot of the three animals together (also 4 seconds). 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
At 17 seconds 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). This sequence ensures that each bumping event involved a dog and a cat and that there was a dog who was the agent of a bumping and a dog who was the patient of a bumping. 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?” (direct object condition).

3.5.3. Test phase
At 15.3 seconds 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, “Which dog did the cat bump?”—direct object 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.
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 2005) 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 time line of proportion of looks toward the target for every frame in the test trial. This time line allowed us to look for general trends in looking across the trials.

As we expected looking behavior to be conditioned by the linguistic stimulus (i.e., we expected participants to look more toward the correct character after the linguistic stimulus was uttered, rather than during the utterance or before it), we want to analyze particular critical time windows. We were able to do this by averaging the proportion of participants looking toward the target for a certain duration of time. We used this method to look at the average proportion of looks toward the target animal in a 1-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, where we expect looking to be conditioned by the linguistic stimuli, that we will be comparing in the following.

Four coders coded these data. Intercoder reliability was always above 90%, and Cohen’s Kappa ≥ 90%.

3.7. Results

By constructing the time lines discussed earlier for every condition and by averaging the proportions of looks toward the target over the critical time windows, we were able to carefully examine data across conditions. In no condition did we find systematic effects of sex of infant, vocabulary level of infant, individual verbs, or order of presentation, so these factors are not included in the analyses we report here.

3.7.1. Selection of data for analysis

Broadly, there are two ways that preferential looking data can be analyzed: by averaging looks to target across a whole trial or by choosing regions of interest where a response to a linguistic stimulus is expected and looking at average looks in windows following these points in time. Seidl, Hollich, & Jusczyk (2003), with relatively simple test trials, were able to use the first method. As the test portion of each of our trials involved three iterations of the target utterance (wh-question), as well as attention-getting audio (“Now look!” and “Can you find him?”), we found it prudent to take the second approach. This gave us three 1-second windows to examine (one following the each iteration of the target question). While the exact time course of apparent question comprehension varied across conditions and age groups, in conditions where we saw time-course evidence of comprehension in any of the critical windows, we consistently found it in the window following the offset of the second target question. This is evident when viewing the looking time lines of the trials: While subjects’ attention shifts between target and distractor, if they ever reliably look at the target, they do so during the 1-second window following the second iteration of the test question. This does not mean we didn’t find

Our choice of a 1-second window for analysis was based on how we designed the stimuli, with a 1-second period after the offset of the question before either the screen went black, further linguistic stimulus was uttered, or the test phase ended. Upon analysis and consideration of data, we determined that such a window was sufficient but not optimal for capturing the time course of processing these questions.
this evidence elsewhere (a glance at the time lines will show that infants looked preferentially at the
target at other points in the test trials as well), merely that for a fair comparison of conditions and
groups, we chose the window where the representative effect was most consistent. Thus while infants
are looking back and forth at the two images throughout the test trial, we look for a consistent pattern.
That is, we look for regions where responses are consistent relative to a question.

Recall that in designing the experiment we weren’t certain whether the infant subjects would need
to hear the test question more than once and whether it would be helpful for them to hear the
question when no distracting visual information was present. Looking at the data, it appears as
though infants do need to hear the question more than once and that the blank screen during the
utterance of the second question ensures that looks are influenced only by the question and not by
features of the video that might attract attention independent of the child’s sentence understanding.
The averages over this region are the data discussed in the following.

### 3.7.2. 15-month-olds

Figure 1 shows the timeline of the entire trial, averaged across all subjects and all trials.

The points that make up the lines represent the average proportion of looks per frame towards the
character who had been an agent in the familiarization events, henceforth “the agent.” That is, we
calculated as our dependent measure the looks toward the agent divided by the sum of looks toward
either the agent or patient, as is standard in reporting proportional looking data in the infant
literature (Waxman et al. 2009). The agentive character is the target response in the subject
conditions and the nontarget response in the object conditions. Looks that are not to the agent
are to the patient, the target in the direct object condition, as frames when the children were not
looking have been excluded from this analysis. The vertical bars delimit a 1-second window
following the offset of each utterance of the target question. Averages by condition from these
critical windows constitute the data analyzed. A 2 × 3 repeated measures ANOVA (condition*ques-
tion) revealed a significant effect of question, $F(2,56) = 3.95, p = .025$, but no effect of condition or
any interaction in the 1-second windows following each question. However, recall that one potential
problem raised earlier with respect to the Seidl, Hollich & Jusczyk (2003) study was that the small
number of trials may have masked participants’ abilities. To investigate this hypothesis, we divided
the data into two blocks, comparing performance in the first three trials with performance in the last
three. Figure 2 shows time lines of the data by block.

After splitting the data into blocks, we performed a 2 × 2 repeated measures ANOVA (con-
dition*block) for the window following each question. In the windows following both the first and
third questions, there were no significant effects or interactions. In the window following the second
question, there was a significant interaction, $F(1,28) = 9.27, p = .005$. At first glance this effect

![Figure 1. Timeline showing mean proportion looks to agent in WH-Questions by 15-month-olds across all subjects and all trials.](image)
appears to be in part driven by a backwards looking pattern in the first block of trials, in which participants in the subject condition look at the patient and those in the object condition look at the agent. However, while the conditions do diverge, whatever is driving this does not appear to be contingent on the linguistic stimuli. Looking at the time line, we can see that the conditions diverge in this way from the very beginning of the trial, before any linguistic stimulus has been uttered. In the second block, the divergences between conditions do appear to be contingent on the linguistic stimuli. Planned comparisons show no significant difference between the two conditions in the 1-second window following the second question in the first block but do show one in the same window in the second block ($t = 2.21$, $p = .035$).

3.7.3. 20-Month-Olds: WH all trials
We analyzed 20-month-olds’ data in the same way as the 15-month-olds’. Figure 3 shows a time line of looks averaged across all subjects and all trials.

A 2 × 3 repeated measures ANOVA (condition*question) for each 1-second window following each question revealed no effects of condition or interactions but did reveal a significant effect of question, $F(2,56) = 3.95$, $p = .025$. As with the 15-month-olds’ data, we split the data into two blocks corresponding to the first three and last three trials (Figure 4).

As with the 15-month-olds, we see some divergences by condition that appear to go in the opposite direction than we would predict during the first block. However, this divergence does not appear to be contingent on the linguistic stimuli, since it begins before the target utterance. During the second block, we do see differences by condition that appear to be contingent on the linguistic stimuli. A set of 2 × 2 repeated measure ANOVAs (condition*block) revealed no significant effects or interactions for either the first or third question but did reveal a significant interaction after the second question, $F(1,28) = 12.01$, $p = .002$. Recall that this is the same window where we saw the same interaction for the 15-month-olds.

---

It is possible that differences between conditions in the audio portion of the action familiarization phase is responsible for this asymmetry in the first three trials, though further research would be required to identify the precise nature of this effect.
3.7.4. Discussion of results

While the effects we uncovered were by no means robust across all of the data, we found a similar pattern in both age groups at a similar point both in the experiment (in the second block) and in the trial (after the second question). We predicted that infants would need more trials, thus expected that if they would succeed anywhere, they would do so in the second half of the experiment. As mentioned, when designing the experiment we did not know how many iterations of a question infants would need to hear nor whether or not the blank screen would aid their comprehension. It looks as though the window following the second question is a sort of sweet spot—due to the repetition, the blank screen during the question, or a combination of both.

This pattern of results suggests that 15-month-olds correctly understood both subject and object *wh*-questions. This suggests, in turn, that the concerns cited with the methodology in the Seidl, Hollich & Jusczyk (2003) article were responsible for the subject-object asymmetry in 15-
month-olds’ comprehension in that work. Crucially, children’s overall success is only evident when looking at later trials. This strengthens the argument that the small number of trials in the Seidl, Hollich & Jusczyk (2003) study did not give 15-month-olds the opportunity to fully exhibit their comprehension abilities. As predicted, 20-month-olds behaved as though they understood both subject and object wh-questions.

These results are compatible with three possibilities: (a) that the adult-like grammar is in place by 15 months, and there is no change in this grammar by 20 months; (b) that the adult-like grammar is not in place in either 15- or 20-month-olds and that all children succeeded by leveraging their partial knowledge of the grammar; or (c) that the 20-month-olds have the correct grammar, and the 15-month-olds succeed by using partial knowledge. Exploring the source of children’s success with wh-questions at 15 and 20 months is the goal of 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, Hollich & Jusczyk (2003), the question remains as to whether 15-month-olds use adult-like knowledge to arrive at the correct interpretation in our task. To probe this question, and to better understand the developmental trajectory of filler-gap dependency resolution, we examined the comprehension of an arguably more difficult filler-gap dependency, the relative clause. We used the same methodology as in Experiment 1, which did not elicit a difference between 15- and 20-month-olds 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

Recall that we have three possibilities: (a) both age groups have adult-like knowledge; (b) both ages groups rely on only partial knowledge; and (c) 15-month-olds rely on partial knowledge, but 20-month-olds have and use adult-like grammatical knowledge.

Before we can fully determine the predictions of these hypotheses, we must make a concrete proposal about how children who do not yet have knowledge of the grammar of filler-gap dependencies in English would use what partial knowledge they do have to arrive at interpretations for these sentences.

By 15 months, children are just beginning to show knowledge of argument structure (Golinkoff et al. 1995; Lidz, White & Baier 2013). This knowledge could support the interpretation of sentences involving filler-gap dependencies. If children know the transitivity of a verb, then they could 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) but without creating a link between the gap and the wh-phrase. 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. This strategy would allow for correct interpretation without forming the link between the filler and the gap and would not require children to parse or interpret the filler to arrive at the correct interpretation. Note also that in Experiment 1, children heard the verb several times during the familiarization phase so that the argument structure of the verb is highly activated by the time of the test phase (cf. Yuan & Fisher 2009; Arunachalam & Waxman 2010).

Children using such partial knowledge would be expected to apply it equally in wh-questions and relative clauses, so we would expect parallel performance across these two constructions. Thus, we
expect no differences between *wh*-questions and relative clauses in children using this partial knowledge.

Children with knowledge of the grammar of filler-gap dependencies would be expected to apply it in both *wh*-questions and relatives. However, as noted earlier, it may be more difficult to identify the filler in a relative clause than in a *wh*-question. Thus, children with the correct grammar might (a) show correct interpretations for both *wh*-questions and relatives or (b) show correct performance only for *wh*-questions.

So, if children have acquired the correct grammar by 15 months, we would expect either that they succeed on both *wh*-questions and relative clauses or that they succeed on *wh*-questions but not relative clauses. Moreover, we would expect either equal or improved performance in 20-month-olds. On the other hand, if children are using an interpretive heuristic at both 15 and 20 months, we would expect them to show parallel performance at both ages and across both constructions. Finally, if children at 15, but not 20, months are using only partial knowledge, then we might expect only the younger group to succeed at relative clauses, with the older children failing only on relatives because of difficulties in identifying the filler.

### 4.3. Participants

Thirty-two 15-month-olds (16 males) with a mean age of 14.27 (range: 14.04 to 15.17) and 32 20-month-olds (16 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, Maryland, area and were acquiring English as native language. Parents completed the MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al. 1993). The 15-month-olds’ mean production CDI vocabulary was (24.7) (range: 0 to 190, out of a total possible 655), and the 20-month-olds’ mean production CDI vocabulary was (107) (range: 9 to 381, out of a total possible 655). While mean CDI scores differed numerically from those found in Experiment 1, the difference was not significant for either age group (two sample t-tests for each of the age group pairs had p values of .33 and .58 respectively). We analyzed the data of infants who completed at least 4 out of 6 test trials (63/64 infants analyzed watched 6/6 test trials), and the trials where the infant was looking at least 20% of the time (this excluded 3 trials). 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, except that the target utterances contained relative clauses rather than *wh*-questions. It is possible that in Experiment 1 the content of the trial and timing of questions and other auditory material may not have allowed subjects sufficient time to comprehend the question before being presented with further auditory stimuli. Such complications could have added more noise to an already difficult task, obscuring the subject’s performance. In Experiment 2 we lengthened the test trial to give subjects more time following the first and third utterances of the target question.

### 4.5. Results

The results of Experiment 2 were analyzed in exactly the same way as those of Experiment 1.

#### 4.5.1. 15-month-olds

As in Experiment 1, we’ll begin by examining the timeline of the entire trial, averaged across all subjects and all trials (Figure 5). The windows demarcated on the timelines reflect 1.5 seconds, as opposed to the one second windows shown with the *wh*-questions. This is to reflect the longer time we gave the children to
respond in the experiment. For consistency with Experiment 1, the analyses only consider 1-second windows; however, the results only get stronger when the 1.5-second windows are analyzed.

A repeated measures 2 x 3 ANOVA (condition*question) for the 1-second windows following each question type revealed no significant effects and no interactions. As in Experiment 1, we split up the data into two blocks to examine data from the first three trials versus the last three (Figure 6).

A set of repeated measures 2 x 2 ANOVAs (condition*block) for each of the target windows revealed no significant effects or interactions in the windows following the first and third questions. In the window following the second question there was both a main effect of block, $F(1,29) = 4.38$, $p = .045$, and an interaction, $F(1,29) = 4.51$, $p = .042$. Note that this is the same place in both the experiment (second block) and trial (second question) where we found evidence of comprehension of wh-questions.

![Figure 5. Timeline showing mean proportion of looks to agent in Relative Clauses by 15-month-olds averaged across all subjects and trials.](image)

![Figure 6. Timeline showing mean proportion of looks to agent in Relative Clauses by 15-month-olds in first versus second block of trials.](image)
Because it appears that the 15-month-olds can comprehend both subject and object relative clauses, these results are consistent with the possibility that 15-month-olds have adultlike knowledge of both *wh*-questions and relative clauses and also with the possibility that they are using partial knowledge to parse. Moreover, they further confirm our hypothesis that the subject-object asymmetry found by Seidl, Hollich & Jusczyk (2003) stemmed from issues in the task itself. As in Experiment 1, we find no subject-object asymmetry, and we see an improvement in performance in later trials as compared to early trials.

In order to determine whether 15-month-olds’ performance derives from an adultlike grammar, we next examine 20-month-olds performance with relatives, as described in section 4.2. To the degree that 20-month-olds show difficulty where 15-month-olds do not, this will provide evidence about whether 15-month-olds’ success derives from adultlike sentence understanding mechanisms.

### 4.5.2. 20-month-olds

As with the 15-month-olds’ data, we will begin with an analysis of all trials.

A 2 x 3 repeated measures ANOVA (condition*question) revealed no significant effects or interactions. We once again divided the data into two blocks (Figure 7).

A set of 2 x 2 repeated measures ANOVAs (condition*block) revealed no effects of condition or block and no interaction in the windows following any of the target utterances. Thus we find no evidence that 20-month-olds can comprehend relative clauses in our task.

### 4.6. Comparison of results between Experiment 1 and Experiment 2

In order to examine whether we really see a difference in performance between 15- and 20-month-olds, and whether this difference in fact hinges on the construction, we built a series of candidate linear mixed effects models to analyze the data from both experiments and both age groups together. As earlier, we focused on the 1-second window following the second question, where the effect was consistently significant. These models explore the effect of the construction being tested (*WH*-Question vs. Relative Clause) and condition (Subject vs. Object) on infants’ looking times. Models were fit in R with the `lmer` function from the `lme4` library (Bates 2005; Bates and Sarkar 2007) using maximum likelihood. The models were then compared using the `anova` function in order to determine whether adding factors explained significant additional variance (Baayen 2008). The set of models that we compared are given in Table 2 (the same set was used for each age group). Model 1 considers only the effect of construction. Model 2 adds a term for the effect of the condition.
independent of construction. Model 3 includes both of these effects and an interaction term. All models included random intercepts for both subject and item.

For 15-month-olds, the analysis of variance comparing these models indicates that M2 is more explanatory than M1 (\( \chi^2 = 5.40, p < .03 \)) but that M3 is no more explanatory than M2 (\( \chi^2 = 0.686, p = .41 \)). This result supports the claim that infants hearing both construction types diverge reliably due to condition but that the construction itself doesn’t affect their comprehension. For 20-month-olds we see a different pattern, where M2 is no more explanatory than M1 (\( \chi^2 = 2.00, p = .15 \)), but only M3, with the interaction between condition and extraction is significantly more explanatory than the other models (\( \chi^2 = 3.63, p = .05 \)). These results further support our claim that 20-month-olds only succeed with wh-questions and really do show different looking patterns when hearing relative clauses.

4.7. Discussion of results

In the relative clause condition, we found a discontinuity in which 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. This decline is unexpected if 15- and 20-month-olds both used the same knowledge. If both groups were adultlike in their knowledge, we would not expect a decline in performance in 20-month-olds relative to the younger 15-month-olds. If both groups used partial knowledge (knowledge of argument structure but not syntax) to parse, we would not expect the use of this knowledge to cause degraded performance only in relative clauses and only in 20-month-olds.

This discontinuous pattern suggests a décollage between the development of knowledge and the necessary deployment systems for this knowledge between 15 and 20 months of age. Our data are most consistent with a developmental trajectory with two properties. First, 15-month-olds succeed at the task using only partial knowledge. Second, 20-month-olds have acquired the grammar of filler-gap dependencies but show difficulties deploying that grammatical knowledge in relative clauses. On this analysis, 20-month-olds’ difficulty with relative clauses derives in part from the difficulty in identifying the filler and retrieving it from memory in a relative as opposed to a wh-question.

At this point, it is reasonable to worry that the link between performance on our task and the knowledge that generates that performance is not straightforward. Of the three groups that devote more looking time to the target, only one (the 20-month-olds in the wh-condition) is argued to succeed because of correct grammatical knowledge. The 15-month-olds’ success is argued to derive from their partial knowledge of the grammar (argument structure but no filler-gap dependencies). And the group that fails at the task (the 20-month-olds in the relative clause condition) is argued to have full grammatical knowledge, but performance factors associated with deploying that knowledge cause them to fail at the task. Perhaps a more straightforward view would be that all children who succeed at the task do so because they have acquired the grammar, whereas those who fail display a performance limitation that derives from something that is only available at the later age. We are sympathetic to this concern.

One possibility, raised by an anonymous reviewer, is that 20-month-olds understand something about the pragmatics of relative clauses that 15-month-olds do not and that this pragmatic knowledge interferes with performance only in the older group. Under this possibility, the grammar of filler-gap dependencies is in place by 15 months, but growing pragmatic sophistication masks this knowledge in 20-month-olds. Perhaps 20-month-olds understand that relative clauses pick out members of a relevant set of options but have trouble constructing this set in our task. Or perhaps
they find the use of show pragmatically odd in our test question, given that this is not an interactive task. This account predicts that 20-month-olds would fail with any kind of relative.

Our first account, however, that 20-month-olds’ difficulties in the relative clause condition stems from difficulties in identifying the filler and retrieving it from memory, makes the opposite prediction. If we make the dependencies in relative clauses easier to resolve, by making the filler easier to identify and retrieve from memory, then 20-month-olds’ performance should improve. Thus if we were to make fillers easier to identify by using wh-words as the relative pronouns (rather than that), these dependencies should be easier for 20-month-olds to resolve. Experiment 3 looks at the comprehension of relative clauses with wh-relative pronouns in an effort to distinguish between these two possibilities. We will see that performance improves in the wh-relatives, arguing against the pragmatic account of 20-month-olds’ failure in experiment 2.

5. Experiment 3: wh-relatives

5.1. Motivation

In Experiment 3, we set out to determine whether 20-month-olds’ failure with relative clauses stems from pragmatic factors related to relative clause knowledge or from difficulty deploying grammatical knowledge of filler-gap dependencies in relative clause comprehension. By deciding between these two possibilities, we can better characterize the knowledge of 20-month-olds, which in turns allows us to better understand the performance of 15-month-olds and how knowledge of filler-gap dependencies grows during infancy.

As noted earlier, failure with relative clauses could be caused by (a) children lacking that aspect of relative clauses that distinguishes them from wh-questions (e.g., clausal embedding, restrictive modification, the discourse conditions on relativization), or (b) a failure to successfully deploy the filler-gap structure in just this case. The latter possibility does not seem unreasonable when we consider the superficial differences between relative clauses and wh-questions that could make the former more difficult to resolve online. These include the optionality of morphologically marked fillers (i.e., wh-words), the possibly less marked displacement of fillers, the lack of do-support, and the lack of question prosody in relative clauses. To investigate this question, we looked at whether 20-month-olds can comprehend relative clauses when processing demands are reduced by using wh-words for relative pronouns (21):

(21) (a) Show me the dog who bumped the cat.
    (b) Show me the dog who the cat bumped.

5.2. Predictions

If 20-month-olds had trouble with relatives in general (due to pragmatic reasons mentioned earlier), they should perform just as badly with these relatives as with the ones in Experiment 2. If, however, they had difficulty identifying the filler (and subsequently associating it with the gap) in our original relative clause condition, using a wh-relative should make the filler salient, cueing the infant in to the dependency. After all, wh-words always signal the need to resolve a filler-gap dependency, making the presence of a wh-word an unambiguous cue to the presence of a dependency.

5.3. Subjects, materials, design, and procedure

Using the same procedure and material as earlier, we tested thirty-two 20-month-olds (16 males) with a mean age of 20.11 (range: 19.20 to 21.00) on their comprehension of the same items as before, the only difference being that these items had a wh-word for a relative pronoun. Participants were native English-acquiring infants recruited from the greater College Park, Maryland, area. As in Experiments 1 and 2, parents completed the MacArthur-Bates Communicative Development
Inventory (CDI) (Fenson et al. 1993). Participants’ mean production CDI vocabulary was (210), (range: 15 to 568, out of a total possible 655). In this case, the CDI vocabulary does differ significantly from the 20-month-olds in both Experiments 1 and 2, a fact we will return to later. All infants viewed all test trials, and we analyzed data from the trials where the infant was looking at least 20% of the time (this included all three trials).

5.4. Results and discussion

The results of this experiment were analyzed just as the results of the previous experiments. A 2 x 3 repeated measures ANOVA (condition*question) revealed no significant effects of condition or question and no interaction (Figure 8).

When we split the data into two blocks and performed the same set of 2 x 2 repeated measures ANOVAs (condition*block) on the 1-second windows following each question, we found no significant effects or interactions (Figure 9). However, visual inspection of the data reveals an enormous effect in the second block, starting during the second half of the first utterance (at around 4 seconds) and extending for almost 3 seconds. The difference between conditions is significant in the 3,300 ms. window (from ~4s. until ~7s. when the screen goes blank) (t = 3.01, p = .005), and each condition differs significantly from chance (Subject: t = 2.07, p = .05, Direct Object: t = –2.57, p = .02). To allay concerns that we are just choosing this window because it is a significant difference in the direction of our liking (not systematically because it aligns with other windows where we find effects in other conditions), we performed t-tests on four other 3.3-second windows: the first 3.3 seconds of each block (20–120, building in a 20-frame [~670 ms.] picture familiarization buffer), the first 3.3 seconds of each block after the pictures come back (frames 320–420) and the next 3.3 seconds after this (frames 420–520). None of these windows showed any significant effect; the only one that approached significance was the 3.3-second (frames 420–520) window in the second trial, where looking patterns again seem to point toward comprehension. Thus, while the looking pattern in this experiment does not quite parallel that found in Experiments 1 and 2, we take this to be evidence of relative clause comprehension at 20 months.

On their own, these results tentatively suggest that 20-month-olds have the appropriate knowledge to comprehend relativization but have difficulty deploying this knowledge when the filler is
difficult to identify. A direct statistical comparison of this data with that of the 20-month-olds listening to that relatives from Experiment 2 is not straightforward, as (a) the 20-month-olds in Experiment 2 don’t succeed anywhere, thus it isn’t clear whether we should compare the region where 15-month-olds succeed and 20-month-olds succeed in Experiment 1 (after the second question) or the region where 20-month-olds succeed in Experiment 2; and (b) the CDI vocabulary of the 20-month-olds in Experiment 3 differed significantly from that of the 20-month-olds in both Experiments 1 and 2.

It is possible that these two points are related, and the higher median vocabulary of the infants in Experiment 3 contributed to their faster processing in this experiment (cf. Fernald, Perfors & Marchman 2006). However, despite this issues that limit the direct comparison, it does appear clear that 20-month-olds perform differently in Experiments 2 and 3, giving tentative support to the claim that they understand relative clauses when the relative marker is a wh-word.

Figure 9. Timeline showing mean proportion looks to agent in WH-Relatives by 20-month-olds averaged across all subjects and trials.

Figure 10. Timeline showing mean proportion looks to agent in WH-relatives by 20-month-olds in first versus second block of trials.
6. General discussion

We began this work seeking to determine the trajectory of the development of both the grammatical knowledge underlying filler-gap dependencies and ability to deploy this knowledge. This included probing the cause of a reported subject-object asymmetry in the comprehension of wh-questions by 15-month-olds. By improving the methodology used to investigate this question, we found no such asymmetry for any age group or construction, highlighting the contribution of methodology to previous results. Our methodology (a) better engaged infants' attention, (b) made the target utterances felicitous, and (c) used multiple test questions and trials to give infants ample opportunity to comprehend questions and succeed at our task. Future work will continue to refine this method for better access to infants' more complex syntactic representations. Given the high level of noise found overall in our data (as compared to similar looking time data looking at knowledge of binding constraints [Lukyanenko, Conroy & Lidz 2013]), in future work it might be best to lower memory demands during the task by using the progressive (e.g., Which dog is bumping the cat?), and measure looking time to two events, rather than requiring infants to encode and subsequently reaccess thematic role assignment across two events.

Turning back to the knowledge question, we do find differences in performance between 15- and 20-month-olds that uncover an apparent discontinuity in development. Whereas both 15- and 20-month-olds' successfully comprehended wh-questions, only the younger group successfully comprehended relative clauses.7 As we would not expect infants to regress in their linguistic knowledge as they progress through development, we asked what these results reveal about children's grammatical knowledge and the systems that deploy this knowledge (cf. Bever 1982). We suggested that it is possible that what looks like success in the 15-month-olds' performance with both wh-questions and relatives reflects a parsing heuristic based in their knowledge of argument structure. We also suggested that 20-month-olds' failure with relative clauses highlighted a crucial step in successfully moving from using partial knowledge like 15-month-olds to an adultlike system. Our third experiment suggested that 20-month-olds do possess the requisite grammatical knowledge for comprehending relative clauses, by showing success in comprehension when we lessen the demands on the parser, making grammatical knowledge easier to deploy in real time.

Our results are suggestive that the developmental trajectory we propose reflects real properties of grammatical growth, but more experimentation is necessary to better characterize children's knowledge at each stage. The suggestive nature of our data opens up a set of compelling questions, which we articulate in the following. In doing so, we will develop our proposal of a strategy that 15-month-olds could be using to parse with partial knowledge and explore how 15-month-olds would be able to use linguistic information made available by this strategy to eventually acquire adultlike knowledge of filler-gap dependencies.

6.1. How 15-month-olds succeed in our experiment

As argued earlier, it is likely that 15-month-olds have not yet acquired full knowledge of filler-gap dependencies. After all, if 20-month-olds are only just sorting out how to deploy this knowledge, it is reasonable to think that this knowledge was not in place 5 months earlier. But if this is the case, what led children to look at the appropriate animal in response to our target utterances?

We noted earlier that 15-month-olds are just beginning to understand the syntax and semantics of argument structure (Golinkoff et al. 1995; Lidz, White & Baier 2013) and that they could use this knowledge to succeed in our task using the following three components: (a) identification of a verb

7On average, 4% of 15-month-olds and 17% of 20-month-olds in the wh experiment, and 16% of 15-month-olds and 27% of 20-month-olds in the relative clause experiment, were reported to know the verbs we used in the stimuli. This discrepancy in specific verb knowledge does not appear to correlate with the observed pattern of results. It is important to note, however, that the verbs were used several times during the familiarization phase of our experiment, prior to the test phase in which the target utterances were produced.
missing a noun phrase needed to fill a required thematic role, (b) recognition of a gap in the argument structure when an expected syntactic argument fails to occur (possibly primed through hearing this verb used with two arguments several times during the test phase), and (c) search through the discourse context for a referent that could fill out this thematic structure. Importantly, if 15-month-olds are relying on this partial knowledge, 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.

If this use of partial knowledge is a general strategy employed by 15-month-olds, we expect that they were relying on the same strategy in the Seidl, Hollich & Jusczyk (2003) study. Recall, however, that Seidl, Hollich & Jusczyk uncovered a subject-object asymmetry in 15-month-olds’ comprehension. We believe that this finding is compatible with our hypothesis in the following way. If 15-month-olds use this strategy in day-to-day parsing (not just as a strategy in our task), then we expect that they have much more experience deploying it in the context of missing subjects than missing objects. That is, it is our impression that many more contexts involve missing subjects: “predicate” wh-questions (e.g., what’s that?), imperatives, subject-auxiliary drop in yes/no questions (e.g., see that?), than missing objects, and thus these infants would have more practice deploying this strategy to recover possible subjects than objects. This is not to say they cannot deploy it to recover objects, just that it might be more difficult, and that object recovery would be more likely than subject recovery to break down in suboptimal experimental conditions.

To be consistent with the implied rejection of the possibility that 15-month-olds have adultlike knowledge, we must determine why children would ever abandon this strategy if it works as well as it appears to. It is possible that children have some expectations about the grammatical conditions that can license a missing argument. One possibility is overt movement of the type seen in filler-gap dependencies. If 20-month-olds understand the relation between movement and subcategorization, realizing that a verb can sometimes find its arguments in displaced positions in the clause, and they can identify the strings that exhibit movement, extragrammatical heuristics like the one proposed earlier would not be available to parse these sentences because of the grammatical constraint requiring that subcategorized arguments must be syntactically realized (Chomsky 1981). At this point, infants would need access to a new parsing mechanism to deploy their updated knowledge of filler-gap dependencies, and this system would be the adult active filling strategy.

Of course, it’s not enough to say why children would abandon this strategy, we must also outline how they would attain the requisite adultlike knowledge. This is an interesting acquisition puzzle, as children must be able to use the output of the immature parser (the subcategorization information) to recognize that there is a dependency that needs to be resolved (cf. Valian 1990; Fodor 1998a, 1998b). In current work we are exploring exactly how this might work. In broad strokes, we propose that when children discover a missing argument (recognized by a mismatch in the number of arguments present and the number expected given the subcategorization frame of the verb), they consider some constrained set of possible operations that might generate this (e.g., a dependency between a structural position and some other material, a new intransitive use of the verb, a null pronoun, etc.). By comparing properties of the strings that each of these hypothesized operations would generate with the properties of the strings observed in the input, the learner would be able to determine if an operation like an A-bar dependency with overt movement should be used to represent the string and be incorporated into the grammar.

While we do not have precise numbers, a preliminary look at the CHILDES Treebank (Pearl & Sprouse 2013) shows that out of a total 149,828 parsed utterances, 37,022 (25%) are fragments (incomplete sentences). Out of these 37,022 fragments, 10,363 (28%) are missing subjects (e.g., a conservative estimate made by approximated by counting fragments consisting of only a VP), and 24,261 (65%) are NPS or other fragments that don’t contain a VP at all, meaning that at most 2,042 (7%) contain an illicitly dropped object. These figures don’t include imperatives (8,735, 7% of total utterances), which necessarily include null subjects, or predicate wh-questions (conservative estimate 2,812 [2%] of total). Altogether (excluding A-bar dependencies) it looks like upwards of 15% of utterances English-acquiring children encounter lack subjects for one reason or another, while only roughly 2% would lack objects.
6.2. **Summarizing our proposal**

At this point, it is useful to summarize our analysis about what changes when 15-month-olds become 20-month-olds and why 20-month-olds have difficulty in our first relative clause experiment. What follows is what we believe to be the most likely hypothesis, but as mentioned earlier, it by no means exhausts the possible explanations for the patterns found in our data.

6.2.1. **Hypothesis**

(1) 20-month-olds have acquired adultlike knowledge of filler-gap dependencies but have yet to fully control an adultlike deployment system, accounting for their difficulty with relative clauses.

(2) 15-month-olds have a non-adult-like knowledge state that includes knowledge of thematic roles, verb meanings, and event structure, along with a non-adult-like deployment system that uses this knowledge in sentence understanding. The combination of their knowledge and deployment system allows them to comprehend sentences containing filler-gap dependencies in our task.

The knowledge states and deployment systems alluded to in our hypothesis, as well as the progression between them are schematized in the following Figure 11.

7. **Conclusion**

In this article, we have identified a case of discontinuous 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 can only correctly interpret *wh*-question and relative clauses with *wh*-words heading the relatives. 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. At any given point in development, the deployment systems will be constrained by the current state of the grammar.
Hence, what it takes to understand a sentence will vary as a function of grammatical development. Moreover, deficiencies in any of these components can lead to distinct outcomes.

First, if children lack the grammatical knowledge that would allow them to assign an appropriate structure to a given utterance, then they might simply fail to understand the sentence. But at the same time they may be able to combine partial knowledge of the grammar with the extralinguistic context in a way that would allow successful understanding, even if they fail to assign an adultlike syntactic representation.

Second, if children have acquired the appropriate grammatical knowledge, various psycholinguistic factors could interfere with understanding in a way that masks that knowledge (Bloom 1990; Valian 1990; Musolino & Lidz 2006; Lidz et al. 2006; Lidz, White & Baier 2013). For example, difficulty retrieving a word from the lexicon (perhaps because it is low frequency) could interfere with a child’s ability to assign the correct parse.

Finally, even with the right grammatical knowledge and a successful parse, children may have a different understanding of the extralinguistic context or of the felicity conditions on the use of a sentence, leading to errors in understanding (Gualmini et al. 2008; Lewis, Hacquard & Lidz 2013). Thus, care is required in identifying the sources of both successful and erroneous understanding, and more importantly, in mapping out a developmental trajectory for children’s knowledge and their ability to deploy it in real time.

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 may have impoverished grammatical representations for these dependencies and that their deployment systems may be appropriate for those representations. Twenty-month-olds, on the other hand, may have accurate adultlike knowledge but have yet to become effective at deploying that knowledge in real time. By examining the nature of 15- and 20-month-olds’ knowledge and deployment, we can better understand not only when children begin to show adultlike knowledge of filler-gap dependencies but how they arrive at this point.

References


Golinkoff, Roberta; Hirsh-Pasek, Kathy; Mervis, Carolyn; Frawley, William & Parillo, Maria. (1995). Lexical principles can be extended to the acquisition of verbs. In Michael Tomasello & William Merriman (Eds.), *Beyond names for things: Young children’s acquisition of verbs* (pp. 185–222). Hillsdale, NJ: Lawrence Erlbaum Associates.


Lidz, Jeffrey; White, Aaron & Baier, Rebecca. (2013). The role of incremental parsing in syntactically conditioned word learning. Submitted manuscript.


Viau, Joshua; Lidz, Jeffrey & Musolino, Julian. Priming of abstract logical representations in 4-year-olds. *Language Acquisition, 17*(1), pp. 26–50


Yuan, Sylvia; Fisher, Cynthia & Snedeker, Jesse. (under review). Counting the nouns: Simple structural cues to verb meaning. Manuscript under review at *Child Development*.

**APPENDIX: 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)

**Test sentences**

**Experiment 1: wh-questions**

Subject Condition/Direct 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/Direct 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.