U-shaped development in the acquisition of filler-gap dependencies: Evidence from 15- and 20-month olds

Annie Gagliardi, Tara M. 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.

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 phenomena that rely on the same kind of grammatical representations, but that diverge in their deployment processes because of surface differences between them. 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 ___?

[1] We restrict our attention in this paper to dependencies involving arguments.

Among *wh*-questions, there are two differences in surface form 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-auxiliary 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).

(5) *Which dog* do you think ___ bumped the cat?

(6) *Which dog* do you think the cat bumped ___ ?

(7) * Which dog did the man make the claim that __ bumped the cat?

(8) * Which dog did the man make the claim that the cat bumped __?

(9) * How$_j$ did the man wonder [which dog$_i$ __$_i$ bumped the cat __$_j$]?

(10) * How$_j$ did the man wonder [which dog$_i$ the cat bumped __$_i$ __$_j$]?

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

(11) 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 that 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 have 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 on 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 obligatory 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 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.

3.2 Predictions

The predictions for this first experiment are straightforward. Regarding 15-month-olds, if the Methodological issues concerning felicity and engagingness 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 the difficulties introduced by these trial properties are amplified 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, then we should see the asymmetry in the current experiments as well. 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 (16 males) with a mean age of 15;0 (range: 14;14 to 15;18) and 32 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, MD area and were acquiring

English as native language. Parents completed the MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al, 1993). 15-month-olds' mean production CDI-vocabulary was (19.2) (range: 0 to 60, out of a total possible 655), and 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 that 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 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 (ling.umd.edu/labs/acquisition/stimuli/wh_s_bump.mp4).

3.4.2 Auditory Stimuli

The audio portion of the stimuli (as outlined below 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 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). All of the 6 action 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 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 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). 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? (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: Schematic of one entire trial

	Number of Frames	*Video*	*Audio**
Character Familiarization Phase	1;00	Black Screen	none
	4;20	Smiling Baby	4;00 Baby Giggle
	1;00	Black Screen	none
	4;00	White dog	Hey look! It's a white dog
	1;00	Black Screen	none
	4;00	Brown dog	Now look! It's a brown dog
	1;00	Black Screen	none
	4;00	Cat	Now look! It's a cat
	1;00	Black Screen	none
	4;00	All animals	Somebody's gonna bump the cat*
Action Familiarization Phase	1;00	Black Screen	none
	7;00	White dog bumps cat, Cat bumps brown dog	Look what's happening! Do you see it? Wow!
	3;00	Black Screen	Which dog's gonna bump the cat?*
	7;00	White dog bumps cat, Cat bumps brown dog	Look what's happening! Do you see it? Wow!
Test Phase	1;00	Black Screen	none
	6;00	Split Screen: White dog, Brown dog	Now look! Which dog bumped the cat?*
	3;10	Black Screen	Which dog bumped the cat?*
	6;00	Split Screen: White dog, Brown dog	Can you find him? which dog bumped the cat?*

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 duration of time. 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. 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. While the exact time course of apparent question comprehension varied across conditions and age groups we consistently saw time-course evidence of comprehension in the one second window[2] following the offset of the second target question. This window makes sense because the 2nd question is uttered while the screen is blank and so looks are influenced only by the question and not by features of the video that might attract attention independent of their sentence understanding. The averages over this region are used in the discussion below.

3.7.1 15-month-olds

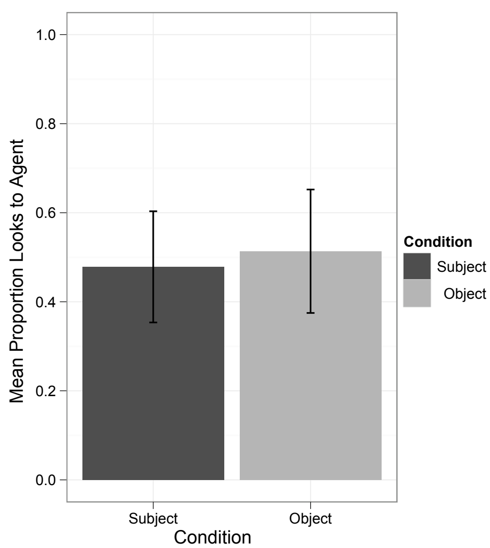

Figure 1: 15-months WH: all trials, 1 second window following 2nd Question

Figure 1 plots looking time during 1-second window following the offset of the 2nd target utterance. The bars represent the average time looking towards the character who had been an

[2] Our choice of a 1 second window for analysis was based on how we designed the stimuli, with a one 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. See Section 5.3 for a discussion of the limitations that such a window introduces.

agent in the familiarization events, henceforth "the agent". This character is the target response in the subject conditions and the nontarget response in the object conditions A one way ANOVA across all trials revealed no effect of condition in the one second windows following any of the questions. However, recall that one potential problem raised above with respect to the Seidl, Hollich and Jusczyk study was that the small number of trials may have masked participants' abilities. Consequently, we also divided the data into two blocks, comparing performance in the first three trials with performance in the last three. Figures 2 and 3 show averages over the window following the second question by block.

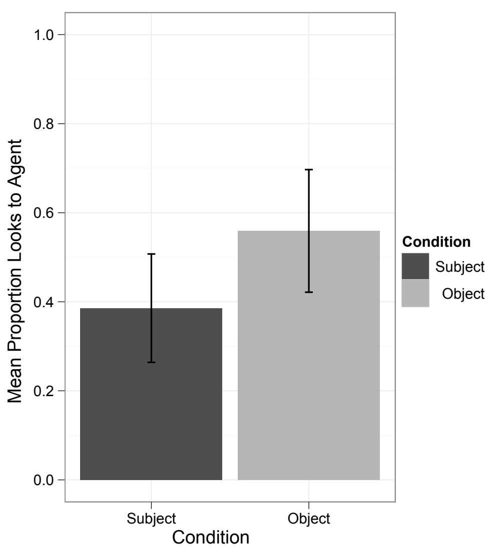

Figure 2: 15-months WH: 1st Block (trials 1-3), 1 second window following 2nd Question

In the first block of trials, a one way ANOVA revealed an effect of condition ($F(1,29) = 4.77$, $p < 0.04$) following the second question. This effect may appear worrisome, as the conditions reliably diverge in the opposite direction from that which we predict based on comprehension of the linguistic stimuli. To better understand the nature of this pattern, we looked at the timeline of looks to the agent in both conditions across the entire trial for the first block of trials. Unlike the effect we will discuss below, this divergence does not appear to be contingent on the linguistic stimuli. That is, it appears before any linguistic stimuli have been uttered and persists across the entire trial. This suggests that whatever is driving this effect is not due to filler-gap dependency comprehension, but some feature of the familiarization influencing the infants' preference to look at certain characters over others.[3]

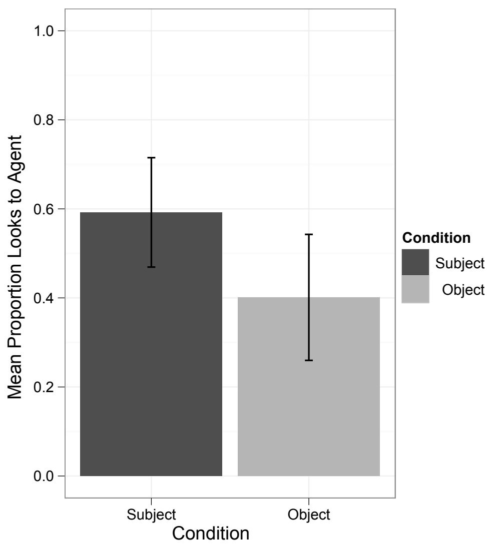

Figure 3: 15-months WH: 2nd Block (trials 4-6), 1 second window following 2nd Question

In the second block, however, it does look as though there are differences contingent on the linguistic information in the windows following questions two and three. That is, these differences emerged in the windows following the offset of the linguistic stimuli, and didn't occur in the beginning of the trial before any stimuli were uttered. A one way ANOVA revealed a significant effect of condition following question 2 ($F(1,29) = 4.72$, $p<0.04$). A 2x2x2 repeated measures ANOVA (condition*block*question) looking at the windows following questions two and three across both blocks revealed a marginally significant interaction between condition and

[3] 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.

block ($F(1,244) = 2.86$, $p<0.10$), and a three way interaction between condition, block and question ($F(1,244) = 4.24$, $p<.05$).

In order to quantify the factors determining looking time in this experiment more precisely, we built a series of candidate linear mixed effects models. As above, we focused on the 1 second window following the second question, where the effect was consistently significant. These models, corresponding to alternate hypotheses about the effect of the block considered (all vs. first block vs. 2nd block) and condition (Subject vs. Object) to infants' looking times were fit in R (R development core team, 2008) 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 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 condition independent of block. Model 3 includes both of these effects and an interaction term. All models included random intercepts for both *subject* and *item*.

Table 2

Model	**Fixed Effects**	**Random Effects**
m1	Block	Subject, Item
m2	Block + Extraction	Subject, Item
m3	Block + Extraction + Block:Extraction	Subject, Item

The analysis of variance comparing these models indicates that m3 is more explanatory than m1 or m2 ($\chi^2 = 64.40$, $p < 1 * 10^{-15}$), further supporting the conclusion that the small number of trials in previous work played a critical role in masking 15-month-olds' ability to understand object questions.

3.7.3 20-month: WH all trials

We analyzed 20-month-olds' data in the same way as the 15-month-olds'. Figure 4 shows the average looking time in the one second window following the second question across all subjects and all trials.

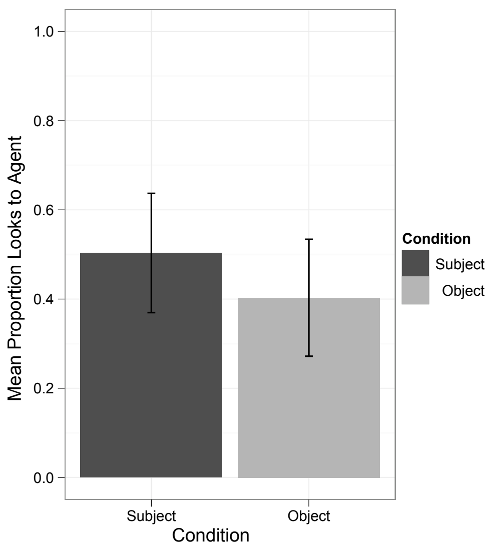

Figure 4: 20-months WH: all trials, 1 second window following 2nd Question

A one way ANOVA for the one second window following the second question revealed no effect of condition. As with the 15-month-olds' data, we split the data into two blocks corresponding to the first three and last three trials (Figures 5 and 6 respectively).

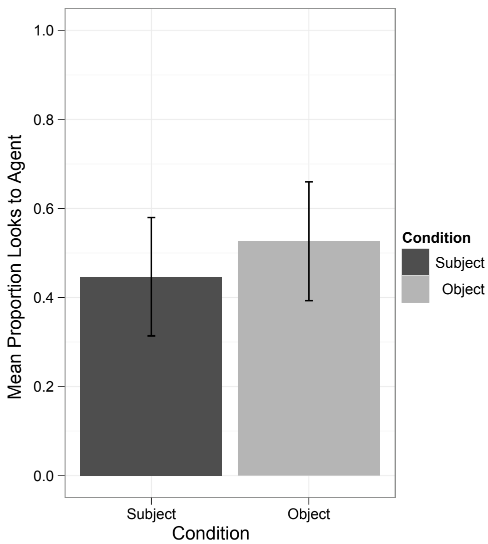

Figure 5: 20-months WH: 1st block (trials 1-3), 1 second window following 2nd Question

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, just as with the 15-month-olds, this divergence does not appear to be contingent on the linguistic stimuli. A one way ANOVA revealed no significant effect of condition.

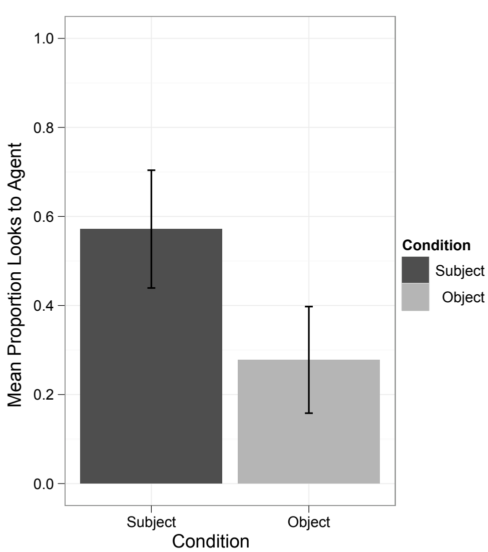

Figure 6: 20-months WH: 2nd Block (trials 4-6), 1 second window following 2nd Question

During the second block, we do see differences by condition that appear to be contigent on the linguistic stimuli. A one way ANOVA revealed a significant difference in the window following the second question ($F(1,29) = 15.8$, $p<0.0005$). A 2x2x2 repeated measures ANOVA (condition*block*question) for data from the second and third question revealed a significant interaction between condition and block ($F(1,244) = 7.5$, $p<0.01$), and no effect of question.

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 2

Model	Fixed Effects	Random Effects
m1	Block	Subject, Item
m2	Block + Extraction	Subject, Item
m3	Block + Extraction + Block:Extraction	Subject, Item

The analysis of variance comparing these models indicates that m3 is more explanatory than m1 or m2 ($\chi^2 = 91.45$, $p < 2 * 10^{-16}$).

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. Crucially, this effect is only evident when looking at second block of trials. This 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.

There are several issues with the content of the trial and timing of questions and other auditory material which could have both not allowed subjects sufficient time to comprehend the question before be presented with further auditory stimuli. Such complications could have added more noise to an already difficult task, obscuring 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 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 persist 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 (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, MD area and were acquiring English as native language. Parents completed the MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al, 1993). 15-month-olds' mean production CDI-vocabulary was (24.7) (range: 0 to 190, out of a total possible 655), and 20-month-olds' mean production CDI-vocabulary was (107) (range: 9 to 381, out of a total possible 655). We analyzed the data of infants that 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. The test phase was 2 seconds longer, due to the reasons mentioned at the end of Section 3.

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 one second window following the 2nd target utterance averaged across all subjects and all trials.

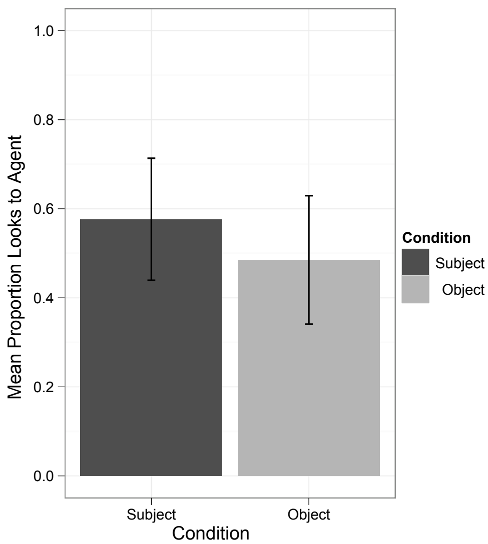

Figure 7: 15-months RC: all trials , 1 second window following 2nd Utterance

A one way ANOVA for the one second window following the second target utterance revealed no effect of condition. As in Experiment 1, we split up the data into two blocks to examine data from the first three trials versus the last three.

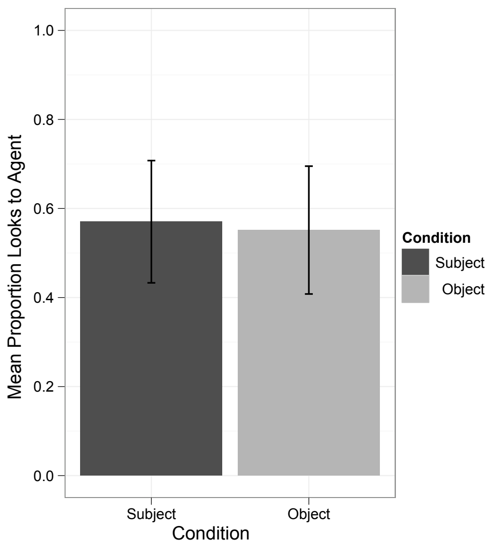

Figure 8: 15-Months RC: 1st Block (trials 1-3), 1 second window following 2nd Utterance

In the first block, a one way ANOVA showed no effect of condition in the critical window.

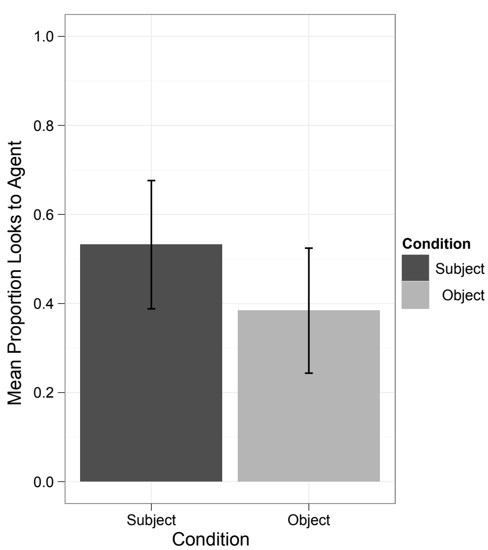

Figure 9: 15-Months RC: 2nd Block (trials 4-6), 1 second window following 2nd Utterance

In the second block, a one way ANOVA revealed a significant effect of condition following the second question ($F(1,29) = 5.71$, $p<0.03$). A 2x2x2 repeated measures ANOVA (condition*block*question) for the windows following the second and third target utterances revealed a marginally significant interaction of condition and block ($F(1,225) = 2.72$, $p = 0.10$) and no effect of utterance.

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

Model	**Fixed Effects**	**Random Effects**
m1	Block	Subject, Item
m2	Block + Extraction	Subject, Item
m3	Block + Extraction + Block:Extraction	Subject, Item

The analysis of variance comparing these models indicates that m3 is more explanatory than m2 or m1 ($\chi^2 = 74.21$, $p < 2 * 10^{-16}$).

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 or 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 more 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 will begin with an analysis of all trials.

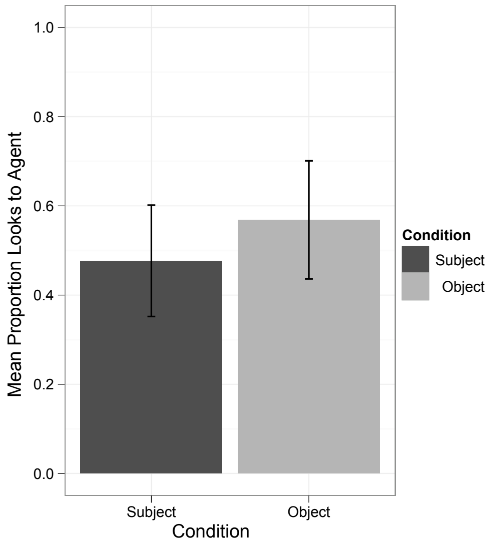

Figure 10: 20-months RC: all trials, 1 second window following 2nd Utterance

A one way ANOVA in the window following the second test utterance showed no significant effect of condition. We once again divided the data into two blocks.

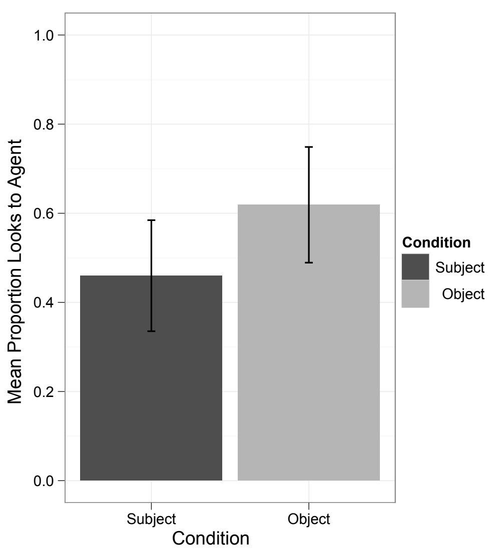

Figure 11: 20-months RC: 1st Block (trials 1-3), 1 second window following 2nd Utterance

In the first block, a one way ANOVA showed a marginally significant effect of condition ($F(1,29) = 3.51$, $p< 0.08$) in the window following the second utterance. It appears however, that this is due to a pattern of switching back and forth across the entire trial that, while it varies by condition it does not appear to be contingent on the linguistic information because it begins well before any linguistic information has been heard.

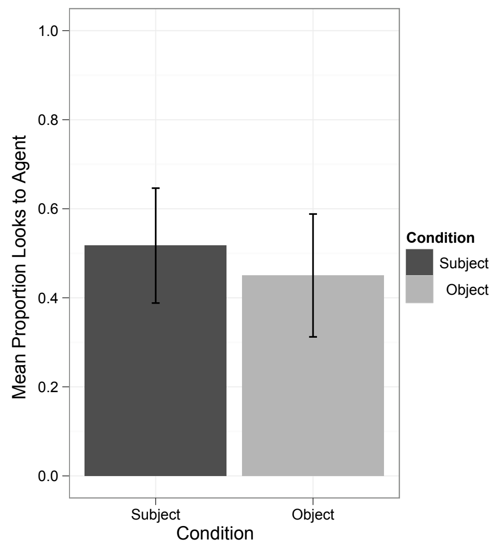

Figure 12: 20-months RC: 2nd Block (trials 4-6), 1 second window following 2nd Utterance

In a one way ANOVA we found no significant effect of condition in the 1 second window following the 2nd utterance (and no effect in the windows following the other two utterances). Similarly, a 2x2x2 repeated measures ANOVA (condition*block*utterance) found no effect of condition, block or question and no interactions.

As in all previous analyses, 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

Model	Fixed Effects	Random Effects
m1	Block	Subject, Item
m2	Block + Extraction	Subject, Item
m3	Block + Extraction + Block:Extraction	Subject, Item

The analysis of variance comparing these models indicates that neither m2 nor m3 are better at explaining the variance than m1 ($\chi^2 = 0.80$, $p = 0.37$ and $\chi^2 = 1.23$, $p = 0.54$ respectively)

4.6 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 in performance from 15- to 20-months is unexpected for two reasons. First, the grammatical parallels between *wh*-questions and relatives leads us to expect that the processing of *wh*-questions and relative clauses should rely on the same mechanisms. Consequently, if children in one age group are able to process one of these constructions, we would expect them to be able to process the other. Thus, 20-month-olds' failure with relative clauses is surprising. 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 décollage between the development of knowledge and the necessary deployment systems for this knowledge between 15 and 20 months of age.

5 General Discussion

We began this work seeking to determine the cause of a reported subject-object asymmetry in the comprehension of *wh*-questions by 15-month-olds. Along the way we improved the methodology used to investigate this question, and found no such asymmetry for any age group or construction, highlighting the contribution of methodology to previous results. Moreover, differences in performance between 15- and 20-month-olds uncover an apparent discontinuity in development. Fifteen-month-olds could comprehend relative clauses whereas 20-month-olds could not.[4] As we would not expect infants to regress in their linguistic knowledge as they progress through development, we must ask what these results reveal about children's grammatical knowledge and the systems that deploy this knowledge. Is it possible that what looks like success in the 15-month-olds' behavior reflects a failure to use adultlike knowledge and deployment systems to parse relative clauses? Could 20-month-olds' failure with relative clauses be highlighting a crucial step in successfully moving from heuristic strategies employed by 15-month-olds to an adultlike system? Below we will explore the implications of these tentative hypotheses.

5.1 Understanding the U-shaped pattern of results

A common view of learning, and one that we will accept here, is that the knowledge, and hence the appropriate deployment system for this knowledge, that is present at one stage will be cumulative across the course of development. That is to say, once children have acquired a given piece of grammatical knowledge, they do not lose this knowledge with subsequent linguistic experience. Our results thus lead us to ask whether it is plausible that 15-month-olds know something about filler-gap dependencies that they subsequently lose by the time they are

[4] [4] On 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.

20-month-olds. The implausibility of such regression across development forces us to examine what other interaction between their developing knowledge and deployment systems could give rise to the patterns of success and failure in our task.

Adopting the position that a child's linguistic knowledge won't regress, we are left with two possibilities to explain the observed discontinuity. First, it could be that 15-month-olds initially acquire a correct characterization of filler-gap dependencies, but some piece of further linguistic knowledge or experience encountered between 15 and 20 months interferes with their ability to use this knowledge. Alternatively, it could be that 15-month-olds haven't yet acquired the requisite knowledge to interpret filler-gap dependencies and instead rely on a temporary heuristic. This heuristic would be rendered insufficient with the acquisition of relevant linguistic knowledge by 20-months.

To determine the plausibility of these two accounts we must carefully outline the knowledge and deployment states children would pass through in each one. By making these possible states explicit we are able to make several predictions that are informing our ongoing work.

5.2. Hypothesis 1: Success means success, and so does failure

The first possibility we consider is that 15-month-olds have acquired adultlike knowledge of filler-gap dependencies and are successfully deploying this knowledge in our task. This would imply that they have both an adultlike knowledge system and a correspondingly adultlike system to deploy this knowledge. Under this hypothesis, 20-month-olds fail not because they lack appropriate knowledge, but because something (knowledge or linguistic experience, potentially in the form of frequency information) is impeding successful online deployment of this knowledge (cf. Lidz 2011). In order for this hypothesis to be viable, there would have to be some linguistic knowledge or experience that could lead to unsuccessful parsing of relative clauses (while leaving parsing of wh-questions intact). It is not immediately clear what this knowledge or experience would be, but further exploration of this question might yield promising results.

5.2. Hypothesis 2: Success means failure, and failure means success.

An alternative possibility is that 15-month-olds have not acquired adultlike knowledge of filler-gap dependencies, and correspondingly lack an adultlike deployment system for this knowledge. Then we must ask, how do they succeed at our task when they fail to have adultlike knowledge and deployment systems? Further we have to ask what is behind 20-month-olds' failure with relative clauses. Have they failed to acquire some crucial piece of knowledge about filler-gap dependencies? Or have they successfully acquired an adultlike knowledge state but can only successfully deploy under certain conditions?

5.2.1 When failure means success

We'll begin by discussing the question of whether 20-month olds fail because they lack the appropriate knowledge to interpret relative clauses or because they lack the appropriate system to deploy this knowledge. Their success with *wh*-questions suggests that they do not lack the requisite knowledge or deployment system to resolve all filler-gap dependencies, so we can narrow down our questions to ask whether they lack knowledge about relative clauses in particular or whether their deployment system is one that only works with *wh*-questions.

Because the structures underlying the filler-gap dependency in *wh*-questions and relative clauses are fundamentally alike, we might take 20-month-olds' success with *wh*-questions as indicative that this common structure is in place. Thus, failure with relative clauses could be caused by children lacking that aspect of relative clauses that distinguishes them from *wh*-questions (e.g., clausal embedding, restrictive modification, the discourse conditions on relativization). Alternatively, the failure with relative clauses could be explained by 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. While we will not present it here, recent results from our lab support this hypothesis, showing that 20-month-olds can comprehend relative clauses when processing demands are reduced, suggesting that they do have the appropriate knowledge for relativization but have difficulty deploying this knowledge (Lidz & Gagliardi, 2010).

Leaving answers to these questions for future work, we are now in a position to consider the following. If 20-month-olds do not lack knowledge of filler-gap dependencies, but do have difficulty deploying this knowledge, why do 15-month-olds do better? What is it that 15-month-olds are doing to perform successfully with both types of filler-gap dependencies?

5.2.2 When success means failure

It could be 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 not unreasonable to think that this knowledge was not intact earlier on. If this is the case, then we must ask if there is any way it would be possible to succeed in our task without knowledge of filler-gap dependencies. That is, are there any other cues, linguistic or otherwise, that could lead a child to look at the appropriate animal in response to our target utterances that don't involve knowledge of filler-gap dependencies or syntactic movement? We believe there may be.

While 15-month-olds may not know about filler-gap dependencies, they may have the rudiments of verb meanings and argument structure in place (Golinkoff et al 1995). Knowing the meaning of a verb implies knowledge of the argument structure and thematic roles associated with it. This knowledge in turn implies knowing that transitive verbs denote events containing two participants. The 15-month-olds' strategy in our task, then, could be a parsing heuristic 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. Note also that in our method, the child hears 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.

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 null argument. One possibility is overt movement of the type seen in filler-gap dependencies. If by 20-months children have the appropriate grammatical structure and constraints to be able to interpret syntactic movement, they would 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 would need access to a new system to deploy their updated knowledge of filler-gap dependencies, and this system would be the adult active filling strategy.

5.3 Formulating a hypothesis to guide future research

In order to guide further investigation in this vein, it will be very useful to formulate a hypothesis based on the possibilities outlined above. What follows is what we believe to be the most likely hypothesis, but as mentioned above it by no means exhausts the possible explanations for the patterns found in our data.

Hypothesis:

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

(ii) 15-month-olds have a non adultlike knowledge state (K_{15}) that includes knowledge of thematic roles, verb meanings and event structure, along with a non adultlike system to deploy this knowledge (D_{15}). The combination of K_{15} and D_{15} allow 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 Figure 13.

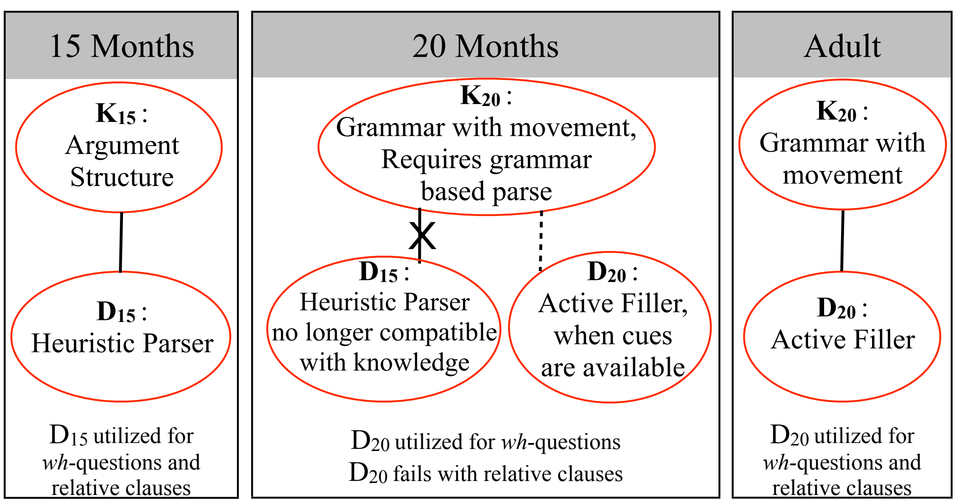

Figure 13: Progression of Knowledge and Deployment from 15 to 20 Months

It is important to recognize that we are not submitting this hypothesis as a claim, as our results cannot support this. This is merely a hypothesis that will drive our future research and, if found to be supported, could account for the patterns of data we have presented in this paper. However, as a hypothesis it works well to make several predictions that are informing are further investigations.

5.2 Predictions

The hypothesis outlined 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 positing 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 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 objects, 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 we are positing that 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. In contrast, if 20-month-olds have adultlike knowledge then they should be sensitive to these extractions. For example, given the scenario outlined above, for an adult the utterance *What did the cat bump and the boy?* (or *what did the cat bump and?)* would be ungrammatical as a violation of the coordinate structure constraint

(Ross 1967). 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 (Lidz & Gagliardi, 2010).

5.3 Limitations

There are several aspects of this data that could be seen as serious concerns. First, the data is very noisy, most likely due to the following factors. The task is complex and subjects could vary greatly in the time course of their responses. Because the analysis requires averaging across trials and across participants, variance in the time course of the responses could cause similar responses that differ in timing to effectively cancel each other out. Additionally, it's possible that we didn't leave enough time to answer the questions, compounded by the fact that there is a significant amount of non-test audio during the test phase, which could potentially alter the course of eye movements as children are processing the target utterances. Finally, the blank screen that occurs between the two halves of the test trial makes it too dark to allow coding. Consequently, if there were predictive eye movements based on the form of the question we would not be able to capture them.

A second issue concerns the backwards looking pattern (i.e., systematic looking at the non-target) that permeates the entire test phase in the first block of trials in the *wh*-question condition, for both 15- and 20-month-olds. This pattern could be due to salience of the target participant in the familiarization and an expectation that the other participant will in turn be highlighted. This does not appear to be a general agent or patient bias, as the agent is preferred in the Object condition, and the patient in the Subject condition. Whatever the precise origins of this curious pattern, it further highlights the utility of using a sufficient number of test trials to allow any task general issue to be filtered out through longer exposure to the task.

While these issues with our data do exist, we nevertheless believe that these data present a compelling picture of filler-gap comprehension at 15- and 20-months. Whatever the problems in the data are, we find consistent patterns conditioned by the linguistic stimuli, and we predict that eliminating some of the more complicated aspects of our test trial would only clarify these results.

5.4 Theoretical Implications

If the hypothesis outlined above proves to be an accurate characterization of the development of filler-gap dependencies, it could also provide 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-months olds would stop using this strategy. If such heuristics were characteristic of mature parsing systems, then we wouldn't expect them to appear early in development and later disappear. 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 would look 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 may not 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 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 adult-like 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

Aoshima, S., Phillips, C., & Weinberg, A. S. (2004). Processing filler–gap dependencies in a head–final language. *Journal of Memory and Language, 51*, 23–54.

Baayen, R.H. (2008). *Analyzing Linguistic Data: A practical guide to statistics using R*. Cambridge, UK: Cambridge University Press.

Bates, D. (2005). Fitting linear mixed models in R. *R News, 5*, 27–30.

Bates, D., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes. R package version 0.99875-6.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 2, 234-272.

Chomsky, N. (1986). *Knowledge of Language*. New York, NY: Praeger

Chomsky, N. (1977). On Wh-Movement. in P. Culicover, T. Wasow and A. Akmajian (Eds.), *Formal Syntax*. New York, NY: Academic Press.

Chomsky, N. (1965). *Aspects of the theory of Syntax*. Cambridge, MA: MIT Press.

Chung, S., & McCloskey, J. (1983). On the interpretation of certain island facts in HPSG. *Linguistic Inquiry, 14*, 704-713.

Clifton, C., Traxler, M.J., Mohamed, M.T., Williams, R.S., Morris, R.K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*, 317–334.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, D. Kartunnen, and A. M. Zwicky (Eds.), *Natural language parsing: psycholinguistic, computational, and theoretical perspectives* (pp. 94-128). Cambridge, UK: Cambridge University Press.

deVilliers, J., & Roeper, T. (1995). Relative Clauses are barriers to wh-movement for young children. *Journal of Child Language, 22*, 389-404.

deVilliers, J., Roeper, T., & Vainikka, A. (1990). The acquisition of long-distance rules. In L. Frazier & J. de Villiers (Eds.), *Language processing and language acquisition* (pp. 257–297). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D., Bates, E., Hartung, J.P., Pethick, S., & Reilly, J.S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. Behavior Research Methods, Instruments, & Computers, 28, 125-127.

Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Fiebach, C., Schlesewsky, M., & Friederici, A. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German wh-questions. *Journal of Memory and Language, 47*, 250–272.

Frazier, L., & Clifton, C. Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 4*, 93-126.

Frazier, L., & deVilliers, J. (Eds). (1990). *Language processing and language acquisition.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Frazier, L., & Flores d'Arcais, G.B. (1989). Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28*, 331-344.

George, L .(1980). *Analogical generalization in natural language syntax*. PhD Dissertation, MIT.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1–76.

Golinkoff, R. M., Hirsh-Pasek, K., Mervis, C. B., Frawley, W., & Parillo, M. (1995). Lexical principles can be extended to the acquisition of verbs. In M. Tomasello & W. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 185-222). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goodluck, H. (in press) Object extraction is not subject to Child Relativized Minimality. Lingua, doi:10.1016/j.lingua.2009.10.005

Gordon P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1411–1423.

Hamburger, H., & Crain, S. (1982). Relative Acquisition. In S. Kuczaj (Ed.), *Language Development: Syntax and Semantics* (pp. 245-274). Lawrence Earlbaum Associates.

Hollich, G. (2005). Supercoder: A program for coding preferential looking (Version 1.5). [Computer Software]. West Lafayette: Purdue University.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29*, 627–645.

Lidz, J. (2011). Grammar-Parser Interactions in the Acquisition of Syntax. In Y. Otsu, ed. *The Proceedings of the Twelfth Tokyo Conference on Psycholinguistics*. Hituzi Syobo Publishing: Tokyo.

Lidz, J. & Gagliardi, A. (2010). Morphosyntactic cues impact filler-gap dependency resolution in 20- and 30-month-olds. (Paper presented at the 2010 Boston University Conference on Language Development).

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 30*, 580-602.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

R development core team. (2008). R: A language and environment for statistical computing. http://www.R-project.org.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Cambridge, UK: Basil Blackwell.

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.

Roeper,T., & de Villiers, J.G. (1994). Lexical links in the Wh-chain. In B.Lust, G. Hermon & J. Kornfilt (Eds) *Syntactic Theory and First Language Acquisition: Cross Linguistic Perspectives Volume II: Binding, dependencies and learnability.* Lawrence Earlbaum.

Ross, J. R. (1967). *Constraints on variables in syntax*. PhD Dissertation, MIT.

Seidl, A., Hollich, G., & Jusczyk, P. (2003). Early understanding of subject and object WH-questions. *Infancy, 4*, 423-436.

Stowe, L. (1986). Parsing wh–constructions: evidence for on–line gap location. *Language and Cognitive Processes, 1*, 227–46.

Stromswold, K. (1995) The Acquisition of Subject and Object Wh Questions. *Language Acquisition, 4*, 5-48.

Sussman, R., & Sedivy, J. (2003). The time course of processing syntactic dependencies. *Language and Cognitive Processes, 18*, 143–63.

Tavakolian, S.L., 1981. The conjoined-clause analysis of relative clauses. In Tavakolian, S.L. (Ed.), *Language Acquisition and Linguistic Theory* (pp. 167-187). Cambridge, MA: MIT Press.

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

Townsend, D., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47*, 69-90.

Traxler, Matthew J., & Martin J. Pickering. (1996). Plausibility and the processing of unbounded dependencies. *Journal of Memory and Language, 35*, 454–475.

Wanner, Eric., & Michael Maratsos. (1978). An ATN approach to comprehension. In Morris Halle, Joan Bresnan, & George A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

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)

Test Sentences
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