RUNNING HEAD: Filler-gap dependency comprehension at 15 months

Filler-gap dependency comprehension at 15 months: The role of vocabulary

Laurel Perkins and Jeffrey Lidz

University of Maryland

Address for correspondence:

Laurel Perkins

Department of Linguistics

University of Maryland, College Park

1401 Marie Mount Hall

College Park, MD 20742

perkinsl@umd.edu

Abstract

15-month-olds behave as if they comprehend filler-gap dependencies such as wh-questions and relative clauses (Gagliardi, Mease, & Lidz, 2016; Seidl, Hollich, & Jusczyk, 2003). On one hypothesis, this success does not reflect adult-like representations, but rather a “gap-driven” interpretation heuristic based on verb knowledge (Gagliardi et al., 2016). Infants who know that *feed* is transitive may notice that a predicted direct object is missing in *Which monkey did the frog feed* __? and then search the display for the animal that got fed. This gap-driven account predicts that 15-month-olds will perform accurately only if they know enough verbs to deploy this interpretation heuristic; therefore, performance should depend on vocabulary. We test this prediction in a preferential looking task and find corroborating evidence: only 15-month-olds with higher vocabulary behave as if they comprehend wh-questions and relative clauses.

1. Introduction

Language acquisition, like learning in general, is incremental: what children can learn about their language depends on what they already know. The way learners experience their input therefore changes over the course of development. Children rely on immature knowledge of their language to identify the structure and meaning of sentences they hear, and the limited portions of the input they can parse become the data they use to draw further inferences about their language. This raises a puzzle. If learners can veridically parse the input, then there is nothing to learn from it; but if they cannot veridically parse the input, then it is unclear how they avoid faulty inferences, or even learn from it at all (Fodor, 1998; Valian, 1990). In this paper, we examine how language learners perceive and learn from their input, given only partial knowledge of their language.

Our case study is the intersection of verb learning and filler-gap dependency acquisition. Learners draw inferences about verbs' argument-taking properties from observing how they distribute in transitive and intransitive clauses (Fisher, Gertner, Scott, & Yuan, 2010; Lidz, White, & Baier, 2017). But it is unclear how those observations of transitivity arise, given the surface variability of transitive and intransitive clauses both within and across languages. For example, recognizing the arguments of a clause may be straightforward for simple declarative sentences like (1), but more difficult for complex clause types that do not follow the language's canonical word order, like the filler-gap dependencies in (2):

(1) a. John ate a sandwich.

b. John ate.

(2) a. *What* did John eat __ ?

b. I made *the sandwich* that John ate __ .

A child familiar with the canonical subject-verb-object word order of English would recognize that sentences like (1a) are transitive, containing both subjects and objects, whereas sentences like (1b) are intransitive, containing only subjects. But recognizing the arguments of the wh-object questions and relative clauses in (2) may be more difficult. These sentences do not have direct objects after the verb, but they are underlyingly transitive. In both cases, an argument acting as the object of the verb (henceforth the “filler,” in italics) stands in a non-local relation to the position in which it is interpreted (henceforth the “gap,” marked by __). In order to identify the structure of these sentences, a child must determine that they do not contain intransitive uses of *eat*, but instead contain filler-gap dependencies– a particular type of non-local predicate-argument relation.

This paper investigates the following question: when do children recognize sentences containing filler-gap dependencies as transitive clauses? Prior literature argues that infants begin to represent the full structure of wh-questions and relative clauses at the age of 20 months, although they behave as if they can comprehend these sentences as early as 15 months (Gagliardi et al., 2016; Seidl et al., 2003). On this hypothesis, acquisition of filler-gap dependencies is driven by developing knowledge of verb argument structure. If a child can detect when a predicted argument of a transitive verb is unexpectedly missing, she may start searching the sentence for signals of the filler-gap dependency that is responsible. Thus, 15-month-olds’ apparent comprehension of these dependencies may not be due to adult-like representations, but rather to a “gap-driven” parsing heuristic based on their initial detection of predicted but missing arguments of verbs.

The gap-driven learning hypothesis predicts that 15-month-olds will only appear to understand wh-questions and relative clauses if they have learned enough verbs to deploy this parsing heuristic. Therefore, performance should depend on vocabulary at this age. In the current study, we confirm this prediction: only high-vocabulary 15-month-olds appear to comprehend filler-gap dependencies. These results point towards a tight connection between verb argument structure and clause structure acquisition in early syntactic development.

2. Background

Filler-gap dependencies are one of the most common types of complex clauses in child-directed speech. English-learning children hear a large number of wh-questions even before their second birthday (around 15% of their total input), the majority of which contain non-canonical word orders (Newport, Gleitman, & Gleitman, 1977; Stromswold, 1995). Relative clauses are somewhat rarer in child-directed speech and are on the surface quite different from wh-questions, but nonetheless share similar underlying structural properties (Chomsky, 1977). Both cases involve a dependency between a displaced argument and the thematic position where the argument is interpreted. The length of this dependency can be relatively short, as in (2), or can hold across arbitrarily long distances, as in (3). Furthermore, neither type of filler-gap dependency can cross certain structures that are "islands" for dependency formation (4) (Chomsky, 1977; Ross, 1967).

(3) a. *What* did Jake believe that Susan claimed that John ate __ ?

b. I made *the sandwich* that Jake believed that Susan claimed that John ate __ .

(4) a. * *What* did Jake believe Susan's claim that John ate __ ?

b. * I made *the sandwich* that Jake believed Susan's claim that John ate __ .

Unlike in English, in some languages wh-questions do not on the surface appear to involve filler-gap dependencies. In “wh-in-situ” languages like Chinese, Japanese, and Korean, wh-phrases are pronounced in their thematic position[1]:

(5) Hufei mai-le shenme (Mandarin Chinese; Cheng, 2003)

Hufei buy-PERF what

‘What did Hufei buy?’

Given the variety of forms that filler-gap dependencies can take, there are several problems that children must solve in order to be able to recognize them in their input. Children need to learn whether they are implicated in constructions like wh-questions and relative clauses and which surface forms in their language signal that they are present. In English, surface signals for filler-gap dependencies include wh-morphology (e.g. words like *what* or *which*), subject-auxiliary inversion, do-support, and relativizers (e.g. *that*, among others). Adult speakers make use of these signals efficiently in sentence processing to identify fillers and predict upcoming gaps (“filler-driven” parsing) (Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier & Clifton, 1989; Frazier & d’Arcais, 1989; Sussman & Sedivy, 2003; Traxler & Pickering, 1996). But because these signals are language-specific, children must learn them. And in order to arrive at the correct interpretation of filler-gap dependencies, children furthermore must identify the particular relation that holds between the displaced argument and a non-local predicate– to identify the thematic position where the filler should be interpreted. This is not a trivial task, as gaps are phonologically null in languages like English.

[1] On many accounts, wh-phrases in these languages still take scope in a higher clausal position by undergoing covert displacement that happens to be inaudible (Aoun, Hornstein, & Sportiche, 1981; Huang, 1982). Some have also argued for non-movement accounts of wh-in-situ, such as binding by a covert operator (Reinhart, 1998), or for different wh-in-situ representations across different languages (Cole & Hermon, 1994). See Cheng (2003) for an overview.

2.1 Hypothesis: Gap-driven learning

On one proposal, children begin to identify filler-gap dependencies by detecting when a phrase stands in relation to a verb that is locally missing a predicted argument (Gagliardi et al., 2016). In other words, although mature parsing of these dependencies is filler-driven, the acquisition of these dependencies may be gap-driven. For example, learners might use the knowledge that some verbs cannot freely alternate between transitive and intransitive frames:

(6) a. Amy fixed her bicycle.

b. *Amy fixed.

A child who knows that *fix* requires a direct object might detect that it is missing after the verb in a wh-object question:

(7) *What* did Amy fix __ ?

She may then be driven to examine the sentence for cues to what happened to this unexpectedly missing argument, and thus start learning the signals of the filler-gap dependency that is responsible: for example, that *what* is an argument wh-word and that do-support can occur in wh-object questions. The detection of a direct object gap will also allow the child to correctly interpret this particular filler-gap dependency, interpreting *what* as questioning some unknown patient of fixing by relating it to the position where *fix* assigns that thematic role.

The gap-driven learning hypothesis proposes a tight relationship between filler-gap dependency and verb argument structure acquisition. Learners need to know which verbs require direct objects, in order to notice when those arguments are needed and missing. In other words, learners need to detect that sentences like (7) contain direct object gaps, rather than intransitive uses of verbs– but in order to do so, they must know which verbs are transitive. This account therefore posits that learners use verb transitivity knowledge to drive filler-gap dependency acquisition, rather than the other way around.

This proposal stands in contrast to another logical alternative: that filler-gap dependency acquisition is actually filler-driven. Under this alternative, the first step for the learner is to use distributional evidence to cluster the wh-words in her language into an equivalence class, by tracking function words that appear at clause boundaries in questions. The second step is to label this cluster *as* the set of wh-words in her language, by identifying that these words stand in particular non-local relationships with a predicate. Having done so, the learner might then use the presence of a wh-word to facilitate verb argument structure acquisition: an argument wh-word signals an upcoming argument gap, and enables the learner to differentiate direct object gaps from intransitive uses of verbs. The filler-driven hypothesis therefore posits that verb transitivity acquisition occurs after learners identify at least some of the filler-gap dependencies in their language– those that involve wh-words.

The first step of this alternative may be feasible. Indeed, Mintz, Newport & Bever (2002) found that an algorithm that clustered words only based on their immediately preceding and following sentence environments in child-directed speech was able to cluster the set of English wh-words when tested on the Nina corpus (Suppes, 1974), although it did not appear to identify such a cluster for the Peter corpus (Bloom, 1970) in CHILDES (MacWhinney, 2000)[2]. However, it is less clear how the learner would proceed from identifying this cluster of words to identifying the particular non-local relationships that they participate in, without using verb transitivity information. Because adjunct wh-words (like *when* and *where*) do not predict upcoming argument gaps, a learner cannot use the mere presence of a wh-word to infer that a verb is being

[2] Note that the goal of this computational model was not to identify closed-class categories like wh-words, but rather to use closed-class items to help identify lexical categories like nouns and verbs.

used in an argument-gap construction, rather than intransitively. Instead, the learner would need to rely on the semantics of particular wh-words to determine which of them are questioning unknown arguments of predicates, and which are questioning times, locations, manners, and reasons. This introduces a new puzzle, which is how a child identifies the semantics of these words– particularly, how a child determines which are the argument wh-words, without first knowing whether they stand in relation to an argument gap.

The gap-driven and filler-driven learning hypotheses make different empirical predictions. Under gap-driven learning, because filler-gap dependency acquisition depends on learning verb transitivity, it should come developmentally later. Under filler-driven learning, the reverse is true: filler-gap dependency acquisition facilitates verb transitivity learning, so it should come at the same time or developmentally earlier.

2.2 Prior Experimental Results

Prior experimental work finds results consistent with the prediction of the gap-driven learning hypothesis: filler-gap dependency acquisition appears to come later than verb transitivity knowledge. However, this evidence is indirect: it is based on an observed U-shaped learning trajectory for wh-question and relative clause comprehension in infancy.

Preferential looking tests of filler-gap dependency comprehension find that 15-month-old English learners behave as if they can comprehend both wh-questions and relative clauses, and sometimes out-perform their 20-month-old peers (Gagliardi et al., 2016; Seidl et al., 2003). In Gagliardi et al. (2016), infants were familiarized to videos where e.g. a dog bumped a cat, who then bumped a different dog. At test, infants saw images of both dogs on different sides of the screen, and heard either a subject or object wh-question or relative clause:

(8) a. *Which dog* __ bumped the cat?

b. *Which dog* did the cat bump __ ?

(9) a. Show me *the dog* that __ bumped the cat.

b. Show me *the dog* that the cat bumped __ .

The researchers assessed infants' looking times towards either one of the two dogs as evidence for how they had interpreted the test sentence. 15-month-olds succeeded on both wh-questions and relative clauses, looking at the correct dog for all sentence types in (8-9). Surprisingly, 20-month-olds performed worse: they succeeded only on wh-questions, and did not look above-chance at the right dog for the relative clauses in (9). This unexpected decrease in performance over development appears to be an example of U-shaped learning, in which a developmental change causes infants to perform (temporarily) worse on a task than they did when they were younger.

Gagliardi et. al. argue that infants at 15 and 20 months used different strategies for processing the sentences in their task. On their account, 15-month-olds' success is not due to adult-like representations of the filler-gap dependencies in these sentences. Instead, 15-month-olds succeeded through a parsing heuristic based on developing knowledge of argument structure: specifically, the ability to predict an argument in a particular position and notice when it is missing. Independent evidence exists for developing argument structure knowledge around this age: for example, the 15-month-olds tested by Jin & Fisher (2014) were able to draw inferences about the meaning of a novel verb on the basis of hearing it in a transitive frame, and Lidz, White & Baier (2017) found that high-vocabulary 16-month-olds predicted an upcoming direct object for a known transitive verb during online sentence processing. In Gagliardi et al's task, infants who had learned that *bump* is transitive, either by prior experience or because it was introduced in transitive clauses during familiarization, may have noticed that a predicted argument is missing in (8-9). They may then have inferred that the experimental task was to locate the referent of that missing argument, by searching the display for the animal that got bumped. Identifying the answer would thus be possible without recognizing that an earlier phrase stands in relation to the verb– that the filler (*which dog* or *the dog*) should be understood as that missing argument. Thus, 15-month-olds may have succeeded without identifying that these sentences contain filler-gap dependencies.

On this hypothesis, 20-month-olds differ from 15-month-olds in their syntactic development. 20-month-olds may have identified that these sentences contain filler-gap dependencies and attempted to represent the filler as an argument of the verb, but faced difficulty deploying their syntactic knowledge online to resolve those dependencies in relative clauses. Gagliardi et al. argue that this is because the cues for argument displacement are less apparent in relative clauses than in wh-questions. For example, relative clauses lack subject-auxiliary inversion and do-support, and the relativizer *that* is homophonous with other words in the language (such as demonstrative *that*). These weaker cues may make it challenging for learners to encode the filler or retrieve it in memory during online sentence processing, resulting in comprehension difficulty. This account predicts that 20-month-olds should improve on relative clauses if they contain stronger cues to displacement, and the researchers confirmed this prediction in a follow-up study that tested wh-relatives:

(10) a. Show me *the dog* who __ bumped the cat.

b. Show me *the dog* who the cat bumped __ .

Because the relativizer *who* is a wh-word, it more strongly signals the presence of the filler in these sentences, potentially facilitating online processing. 20-month-olds were able to identify the correct dog for these wh-relatives, indicating that processing difficulty may have indeed have been responsible for their earlier failure with *that*-relatives. 15-month-olds, by hypothesis, faced no difficulty with *that*-relatives because they were not attempting to resolve any non-local dependency in these sentences.

Although the evidence is indirect, these results point towards a developmental trajectory of filler-gap dependency acquisition that supports the gap-driven learning hypothesis. These results suggest that the ability to identify filler-gap dependencies emerges around the age of 20 months. Younger infants, such as the 15-month-olds in these studies, may be able to infer what the experimenter is talking about by using their argument structure knowledge– by detecting predicted but unexpectedly missing arguments of transitive verbs, and searching the discourse for their referents. By hypothesis, this gap-driven search is what drives the identification of filler-gap dependencies between the ages of 15 and 20 months: as infants attempt to integrate more of the linguistic material in the sentence into a complete parse, they identify that a displaced argument stands in relation to that gap. Thus, it is consistent that gap detection precedes filler-gap dependency acquisition.

2.3 This paper

In this paper, we aim to provide new evidence consistent with the gap-driven account of filler-gap dependency development. If 15-month-olds do not yet identify filler-gap dependencies in sentences that contain them, but rather use a heuristic based on argument structure knowledge to infer what the experimenter is asking in these sentences, this makes the following prediction. 15-month-olds will appear to comprehend wh-questions and relative clauses only if they have learned enough verbs to reliably identify transitive clause structure and recognize when arguments are unexpectedly missing. We therefore expect vocabulary, as a proxy for verb knowledge, to predict children's performance at this age. In the current study, we test and confirm this predicted effect of vocabulary on 15-month-olds' filler-gap dependency comprehension.

3. Method

We tested 15-month-olds' comprehension of wh-questions and relative clauses using a preferential looking task based on Gagliardi et al. (2016), modified to reduce memory demands. In the prior study, infants saw still images of event participants at test and needed to remember who did what to whom in order to respond to the test sentence. In our task, we presented looped videos of events during the test phase, so infants had access to participant role information while hearing the test sentences. Infants were tested in a 2x2 between-subjects design, crossing sentence type (wh-question vs. relative clause) and gap site (subject vs. object). Sample test sentences are provided in Table 1. To simplify our design, we tested only *that*-relative clauses, which 15-month-olds appeared to comprehend in prior work.

Table 1. Sample test sentences (WH: wh-question, RC: relative clause)

	Subject	***Object***
WH	Which monkey is feeding the frog?	Which monkey is the frog feeding?
RC	Find the monkey that's feeding the frog.	Find the monkey that the frog is feeding.

3.1 Participants

Participants included 64 typically-developing infants (32 males) between the ages of 14;14 and 15;18 (mean: 14;29). Participants were recruited from the greater Washington, DC area and were included in the final sample only if they heard English during at least 80% of their waking hours. We analyzed data from trials in which infants attended for at least 20% of the time during the test questions, and only included infants who had at least 4 out of 6 usable trials. An additional 15 infants were tested but not included in the final sample due to fussiness or inattention (9), equipment malfunction (2), less than 80% English exposure (3), or a diagnosed developmental disorder (1).

Vocabulary was assessed by parental report using the Words and Sentences MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1993). Mean total productive vocabulary was 19.2 words (median: 12.5, range: 0-119). Mean productive verb vocabulary was 0.5 words (median: 0, range: 0-10). Because verb production was extremely low overall, we used total productive vocabulary as a proxy for verb knowledge in our analyses, under the assumption that the two are highly correlated and that production often lags behind comprehension in language development.

3.2 Materials

Videos of animal puppets (a frog and two different-colored monkeys) were filmed against a white background. Videos were digitally edited using Adobe Premiere video editing software to ensure consistent brightness and color contrast. Six events were filmed to depict the six familiar transitive verbs used in Gagliardi et al. (2016): *bump, hug, kiss, feed, tickle*, and *wash*. To facilitate counterbalancing of the participant roles of the monkey puppets, two pairs of videos were created for each event. In one pair, the frog acted on the brown monkey and the black monkey acted on the frog; in the second pair, the frog acted on the black monkey and the brown monkey acted on the frog. The timing of each video was edited to be consistent (6.8s each), and events with shorter durations were looped and presented twice. Fig. 1 shows sample still images from one pair of videos.

Fig. 1. Sample video stimuli, one pair of videos for ‘feeding’ event

Audio stimuli were recorded by a female native speaker of American English using infant-directed speech. The audio was edited and combined with the video in Adobe Premiere. Table 2 shows the full set of audio stimuli during a trial. Where audio was the same across trials, the same recording was used to maintain consistency.

Table 2. Structure of a trial. Asterisks mark audio that differs by condition.

Phase	***Audio***	***Video***
Familiarization	*Look, feeding! Somebody’s feeding somebody.*	Frog feeds black monkey (6.8s)
	Look, more feeding! Somebody’s feeding somebody.	Brown monkey feeds frog (6.8s)
Salience	*Wow, look! They’re different!*	Both videos on loop (6.8s)
Test	*Which monkey is feeding the frog?**	Both videos on loop (6.8s)
	*Which monkey is feeding the frog?**	Blank screen (4s)
	Silence	Both videos on loop (6.8s)

3.3 Procedure

Infants were seated on a parent’s lap or in a highchair in a dimly-lit soundproof room, 6 feet away from a 51” widescreen plasma television. Parents were asked not to talk to their infants or direct their attention, and if they held their infant on their lap, they were asked to look away from the screen or wore a visor to prevent them from seeing the screen. A camera located above the television was used to video-record each experiment at a capture rate of 29.97 frames per second. The camera's pan and zoom were controlled by an experimenter watching on a monitor in another room, to ensure that infants' faces stayed within the frame throughout the experiment. Each experiment lasted 5.2 minutes.

The structure of the experiment is as follows. Infants were first introduced to the three animal puppets: the frog and the two monkeys. Each puppet appeared on the screen for 7 seconds with accompanying audio naming the animal (e.g. *Look, it's a brown monkey! Do you see the brown monkey?*). The two monkeys were differentiated by fur color (brown versus black) to ensure that infants perceived them as distinct.

After character introductions, each infant saw 6 trials consisting of a familiarization phase, a salience phase, and a test phase (Table 2). During the familiarization phase, infants saw a pair of videos for a particular event. In one video, the frog was the agent of the event and one of the monkeys was the patient; in the other video, the second monkey was the agent and the frog was the patient. The two videos were separated by a 0.5 sec blank gray screen. The order of the two videos was counterbalanced across trials, such that the first video showed the frog as the agent for half of the trials and a monkey as the agent for the other half of the trials. During each video, infants heard audio labelling the event in a transitive frame that did not specify the identity of either of the two participants (e.g. *Look, feeding! Somebody's feeding somebody*). This differed from the familiarization audio presented in Gagliardi et al. (2016), which named both participants in a way that varied by condition: in that study, a version of the test sentence for each condition was previewed during familiarization (e.g. *Which dog is gonna bump the* *cat?*), raising the possibility that infants were already differentially biased to look towards one of the two animals prior to the test phase. Our modified familiarization audio was intended to eliminate this confound by keeping the familiarization phase consistent across conditions.

The two familiarization videos were then played simultaneously on loop in a smaller size, on different sides of the screen. The actions moved towards the edge of the screen in order to draw infants' eyes away from the center. The left-right position of the target video for each condition was counterbalanced across trials. During the salience phase, infants heard uninformative audio (e.g. *Hey, look! They're different!*), allowing them to adjust to the split-screen format and examine both videos before the test phase began. The salience phase was also intended to reveal any baseline preferences for one of the two videos.

The test phase began after 6.8 seconds. The two looped videos continued to play as infants heard a test sentence that varied by condition (e.g. *Which monkey is feeding the frog?* or *Which monkey is the frog feeding*?). After another 6.8 seconds, the screen went blank and infants heard a second repetition of the same test sentence, to allow them to parse it without visual information. The videos re-appeared after 4 seconds and played silently for another 6.8 seconds until the end of the trial. The decision to repeat the test sentence twice was motivated by the timing of the effects found in Gagliardi et al. (2016), which generally did not emerge until after the second presentation of the test sentence. It is plausible that infants needed a couple of opportunities to parse the sentence in order to demonstrate their comprehension in their task. However, as our task differs substantially from the prior study in both structure and timing (our test phase is much longer), we could not predict *a priori* that the timecourse of infants' behavior would be the same.

Within each condition, infants were assigned to one of two lists to counterbalance the participant roles of the two monkeys across events: in one list, videos for a particular event showed the black monkey as the agent and the brown monkey as the patient, and in the second list, these roles were reversed. Infants were also assigned to one of two random trial orders for each list. To focus infants' attention, trials were interleaved with either a 4-second still image of a baby face with audio of a baby giggling, or 14-second video of moving toys accompanied by music. Infants were given a break if they became restless or fussy.

4. Results

4.1 Data Preparation

The videotaped recordings of the salience and test phases for each trial were coded offline by an experimenter blind to the infant's experimental condition. Experimenters used Supercoder (Hollich, 2005) to advance each muted video frame by frame and code whether infants were looking to the left or right of the screen, or neither. Two coders coded these data, and intercoder reliability was established to be above 90% (Cohen's Kappa > 0.90).

4.2 Overall analysis

We first examined the general timecourse of looking patterns within the salience and test phase of a trial. At each frame, looks to the target and the distractor videos were averaged across all trials for a participant, and across all participants for a condition. This provided a timecourse of the mean proportion of looks towards the target video during a trial, out of looks towards either the target or the distractor.

To determine whether looking patterns were conditioned on the linguistic stimuli, we selected three windows of analysis within the looking timecourse. The selected windows were the three seconds following the offset of each linguistic stimulus: the uninformative audio during the salience phase ("Salience Window"), the first repetition of the test sentence ("Test Window 1"), and the second repetition of the test sentence ("Test Window 2"). Note that these windows of analysis were longer than the 1.5-second windows chosen by Gagliardi et al. (2016), who were limited by the short length of their test phase; in our case, a longer test phase allowed us more time to examine infants' looking patterns.

Fig. 2 displays the overall looking timecourse by sentence type. In these graphs, we plot proportion of looks towards the video in which the monkey was the agent of the event, which was the target video for infants who heard subject-gap sentences (Subj) and the distractor video for infants who heard object-gap sentences (Obj). Looking times are divided into three panels: looking time during the salience phase on the left, during and following the first test sentence in the middle, and following the second test sentence on the right. We do not plot looking times during the second test question, as that question was presented over a blank screen. Instead, we plot looking times starting from the point when the videos re-appeared on the screen until the end of the trial. At first glance, no systematic differences by condition are visually apparent during these timecourses, with the exception of a fairly large backwards-looking pattern during the first test sentence in the wh-question condition: infants appeared to be looking at the wrong answer even before the question began, and this persisted until roughly a second after the question finished.

Fig. 2. Timecourse of mean proportion looks to the monkey agent, all participants

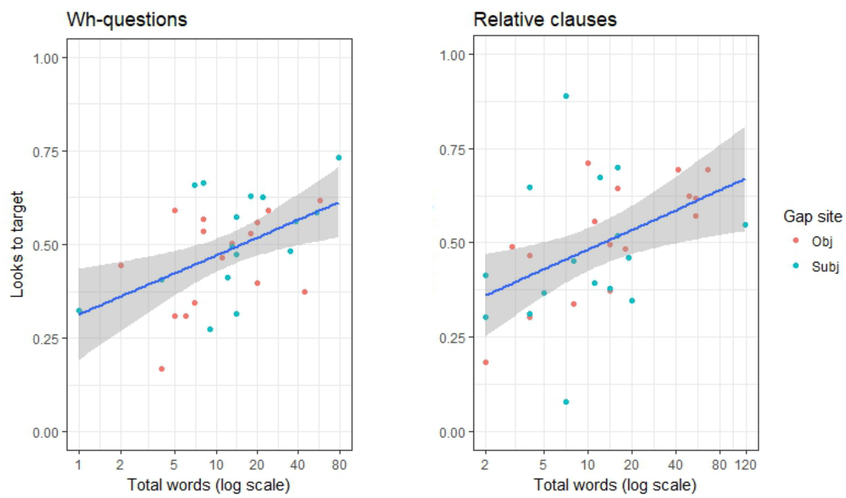
The shaded gray regions represent the three selected windows of analysis, marking the three seconds after the offset of the uninformative salience audio or the first or second test sentences. Looking times during these windows were averaged for analysis. A 2x3 repeated-measures ANOVA (mean looks to monkey agent ~ gap site * window) was conducted for each sentence type (wh-questions and relative clauses), and found no main effects or interactions. Thus, when we average across all participants, we do not find a preference for the target video for any condition during these windows of analysis. This may be the result of (at least) two underlying causes. Either infants in general failed to identify the target video for their condition on our task, or some infants did actually succeed, but their responses were masked by another population of infants who did not. The latter possibility would be expected if infants' performance depended on their vocabulary, as predicted under our hypothesis.

4.3 Vocabulary analysis

To test for the predicted vocabulary effect, we conducted a linear regression predicting proportion looks to the target video based on log-transformed total vocabulary as reported on the MCDI, in each of the three windows of analysis. To control for an effect of age, we also included infants' age in days in the regression model. For both wh-questions and relative clauses, a significant regression equation was found during Test Window 1, corresponding to 3 seconds following the first test question (WH: $F(2, 29) = 4.658$, $p<0.018$, $R^2= 0.243$; RC: $F(2, 29) = 4.053$, $p< 0.028$, $R^2= 0.218$). Vocabulary was a significant predictor for both sentence types (WH: $t= 3.044$, $p< 0.005$; RC: $t= 2.815$, $p< 0.009$), but age in days was not. Fig. 3 displays the relationship between looks to target and vocabulary in this first test window. No significant regression equation was found for the Salience Window or for Test Window 2.

To determine whether gap site (subject vs. object) also affected performance, in addition to vocabulary, we conducted a second linear regression in Test Window 1, adding gap site as a categorical variable. With this more complex model, a significant regression equation was found for wh-questions, and a marginally significant equation was found for relative clauses (WH: $F(3, 28) = 3.364$, $p < 0.032$, $R^2= 0.265$; RC: $F(3, 28) = 2.676$, $p< 0.066$, $R^2= 0.223$). Vocabulary remained a significant predictor for both sentence types (WH: $t= 2.889$, $p< 0.007$; RC: $t= 2.698$, $p< 0.011$), but gap site was not. We therefore find that only vocabulary predicts infants' performance on these sentence types; we do not find a difference in performance on subject vs. object-gap sentences.

Fig. 3. Relationship between vocabulary and looks to target during Test Window 1



We visualized the vocabulary effect by conducting a median split on vocabulary and plotting the timecourse of looks to the monkey agent by both sentence type and vocabulary group. This timecourse is displayed in Fig. 4, with gray bars still marking our windows of analysis. For both wh-questions and relative clauses, looking times for high-vocabulary infants appear to diverge in the expected direction during or shortly after the first test question. This is the window where vocabulary was found to be a significant predictor of performance. High-vocabulary infants who heard a subject-gap sentence appear to look more at the monkey agent, and high-vocabulary infants who heard an object-gap sentence appear to look more at the monkey patient; low-vocabulary infants do not show this same pattern.

Fig. 4. Timecourse of mean proportion looks to the monkey agent by vocabulary group

However, these graphs also reveal that the timecourse of looking is slightly different for high-vocabulary infants who heard wh-questions and relative clauses. For wh-questions, conditions visually diverge in the expected direction roughly one second after the offset of the first test question, preceded by the same pattern of backwards looking that was observed in the overall data. For relative clauses, conditions visually diverge slightly before the offset of the first test sentence, until two seconds after sentence offset. We thus see an effect already in the first half of our window of analysis for relative clauses, but this effect doesn't emerge until the second half of our window for wh-questions. It is unclear why the timecourse of looking for these sentence types would differ, but this means that setting our windows of analysis *a priori* to the three seconds after sentence offset did not accurately capture infants' behavior in both conditions (see Delle Luche, Durrant, Poltrock, & Floccia, 2015 for similar concerns about setting a fixed window of analysis before examining the looking timecourse of an experiment).

To address this issue, we conducted a one-way repeated measures ANOVA comparing looks to target in the first and second halves of Test Window 1. For high-vocabulary infants, a significant effect of window half was found for relative clauses ($F(1, 14) = 11.29$, $p<0.005$), and a marginally significant effect was found for wh-questions ($F(1, 16) = 4.20$, $p<0.057$). No effects were found for low-vocabulary infants. Planned comparisons revealed that high-vocabulary infants looked at the target significantly above chance during the first half of Test Window 1 for relative clauses ($t(14) = 3.23$, $p<0.006$) and during the second half of Test Window 1 for wh-questions ($t(16) = 2.21$, $p<0.042$).

5. Discussion

These results show a predicted effect of vocabulary on 15-month-olds' comprehension of wh-questions and relative clauses. Although the timecourse of sentence processing appears to be different for the two types of sentences, high-vocabulary infants succeeded at identifying the target video during or shortly after the first test sentence for both sentence types. Low-vocabulary infants did not show this same pattern of success.

These results partially replicate the findings in Gagliardi et al. (2016): it appears that 15-month-olds can successfully identify the correct answer for both wh-questions and relative clauses. However, there are some differences in our results compared to prior work. We find that successful performance is modulated by vocabulary, whereas Gagliardi et al. found that successful performance was modulated by the timecourse of the experiment: their effects only showed up during the second block of trials that infants saw. Furthermore, we find effects during the first presentation of the test sentence, whereas Gagliardi et al. found effects during the second sentence presentation. These differences may be due to differences in the design of the current task. Our test phase is substantially longer than that in Gagliardi et al., which may have given infants more time to process the stimuli earlier in the experiment. Presenting looped videos of the events at test, rather than still images of event participants, was also intended to reduce demands on memory and facilitate sentence processing. But processing the test sentence while viewing two simultaneously moving videos on the screen also introduces its own challenges, requiring infants to inhibit attention towards the movement in the distractor video in order to demonstrate a preference for the target video. Thus, further work is needed to determine the extent to which differences in results across the two studies may have been driven by task effects.

Despite these differences, our results are consistent with Gagliardi et al.'s (2016) account of 15-month-olds' success on their task: namely, that 15-month-olds use argument structure knowledge to identify the right answer, without representing the filler-gap dependency in these sentences. Under this argument structure heuristic hypothesis, vocabulary is a proxy for more verb knowledge, resulting in an improved ability to identify transitive clause structure and notice predicted but missing arguments of transitive verbs. Infants may be able to infer that the experimenter is asking about that missing argument, without realizing that the argument has been displaced via wh-movement to a higher clause position.

However, as vocabulary is correlated with many capacities in development, these results are also consistent with other accounts of infants' behavior. Perhaps the high-vocabulary 15-month-olds in our sample have actually acquired the syntax of wh-questions and relative clauses in English, and are arriving at the correct answer in our task not through an interpretation heuristic but through adult-like representations of the filler-gap dependencies in these sentences. Under this first alternative account, vocabulary would be an index of broad syntactic development. On a second alternative account, perhaps all 15-month-olds have acquired the syntax of these sentence types, but only high-vocabulary infants can successfully deploy their knowledge to resolve the filler-gap dependencies in these sentences during online sentence processing. Here, vocabulary would be an index of faster or more automatic parsing abilities.

We find both of these alternatives less plausible given evidence that infants' syntactic abilities are still quite immature at this age. Infants prior to 20 months show variable abilities to recognize certain verbal and nominal morphology (Kouider, Halberda, Wood, & Carey, 2006; Santelmann & Jusczyk, 1998) and map between a word's grammatical category and its meaning (Booth & Waxman, 2009; He & Lidz, 2017). Argument structure knowledge appears to be just emerging at this age, and is apparent mostly for high-vocabulary infants (Lidz et al., 2017) and in simplified task designs (Jin & Fisher, 2014). We therefore find an account based on developing argument structure easier to reconcile with the rest of the literature on infants' syntactic development than an alternative account based on adult-like representations of filler-gap dependencies.

Furthermore, the argument structure heuristic hypothesis provides a more parsimonious account of both the current results and prior work. This hypothesis predicts a specific effect of vocabulary in our task, and can also account for the independently observed relationship between vocabulary and emerging argument structure knowledge found in Lidz et al. (2017). Under the alternative account on which higher-vocabulary infants have acquired the syntax of wh-questions and relative clauses, a vocabulary effect would also be predicted in our task. However, this hypothesis appears to miss a generalization: it does not appeal to the relationship between vocabulary and emerging argument structure knowledge evident around this same age, and thus misses a potential link between the current results and those found in Lidz et al. (2017).

Moreover, under the alternative account on which all infants have acquired wh-question and relative clause syntax, but higher-vocabulary infants have faster or more automatic parsing abilities, a vocabulary effect is merely accommodated. If infants at all vocabulary levels have acquired the syntax of these sentence types, it could just as well have been the case that they would have no difficulty parsing the sentences in our task. Thus, the hypothesis that vocabulary indexes parsing ability can accommodate but does not explain the current results. Furthermore, if the high-vocabulary 15-month-olds succeeded at comprehending both wh-questions and relative clauses in our task due to superior parsing abilities, we lose an account of why 20-month-olds showed poorer performance on relative clauses compared to wh-questions in Gagliardi et al. (2016). Recall that the researchers explained these findings by appealing to mature knowledge of wh-dependencies at 20 months, but difficulties deploying this knowledge online when processing relative clauses. This account received support from the observation that 20-month-olds improved when tested on wh-relatives compared to *that*-relatives. But if the high-vocabulary 15-month-olds in our sample are able not only represent the filler-gap dependencies in *that*-relatives, but also to parse them effectively online, presumably 20-month-olds would likewise have this ability. We thus lose an explanation for the prior finding that infants show different performance on relative clauses at these two ages.

Finally, if infants at any vocabulary level were representing the filler-gap dependency in these sentences, we might expect to see a signature of filler-gap dependency processing: asymmetrical performance on object-gap sentences compared to subject-gap sentences. A substantial literature on preschoolers' relative clause comprehension has found worse performance on object relatives than subject relatives (Adani, Van der Lely, Forgiarini, & Guasti, 2010; Arnon, 2009; de Villiers, Flusberg, Hakuta, & Cohen, 1979; Friedmann, Belletti, & Rizzi, 2009; Goodluck & Tavakolian, 1982; Hamburger & Crain, 1982; Kidd & Bavin, 2002; Labelle, 1990; Sheldon, 1974; Tavakolian, 1981; among many others). Adults also display slower reading times for object compared to subject relatives during online sentence processing, which may be due to how memory is accessed when resolving dependencies over intervening linguistic material (see Wagers & Phillips, 2014, for a review). If difficulty resolving longer dependencies for object relatives is responsible for the subject-object asymmetry, then we predict asymmetrical performance only if children are indeed attempting to resolve the dependency between the filler and the gap in these sentences.

For both wh-questions and relative clauses, the 15-month-olds we tested showed no evidence of asymmetrical performance with subject- vs. object-gap sentences. Although it is difficult to reason about a null effect, the lack of a subject-object asymmetry is consistent with the hypothesis that infants do not yet represent the filler-gap dependencies in these sentences at this age. Furthermore, this lack of an asymmetry reproduces the results of Gagliardi et al. (2016), who also found comparable performance on subject and object wh-questions and relative clauses at 15 months. These more recent findings run contra to the results of an earlier preferential looking study by Seidl et al. (2003), in which 20-month-olds were able to comprehend both wh-subject and object questions, but 15-month-olds were only able to comprehend subject questions. Gagliardi et al. (2016) propose that the asymmetry found by Seidl et al. (2003) may have arisen due to methodological factors, including a within-subjects design with a small number of trials and scenes that made wh-questions less pragmatically felicitous. We believe that the continued lack of a subject-object asymmetry in the current work makes their interpretation of the earlier findings more plausible.

Further work aims to find more direct evidence that 15-month-olds do not yet represent the filler-gap dependency in a wh-question, and that 20-month-olds do. In an ongoing study, we use a listening time task (Maye, Werker, & Gerken, 2002; Shi, Werker, & Cutler, 2006) to test whether infants distinguish auditorily presented filled-gap wh-object questions, such as **Which dog did the cat bump him*?, from questions with gaps. Sentences are presented in the absence of referential context: we no longer measure whether infants can identify an event that matches the sentence, but merely whether infants listen longer to a grammatical or to an ungrammatical sentence. Under our hypothesis, 15-month-olds should process a filled-gap question the same way they would process a simple transitive clause with no wh-phrase, because the argument structure requirements of transitive verb have been locally satisfied and children at this age do not represent *which dog* as an object of the verb. However, older infants should process a filled-gap question differently from a simple transitive clause if they represent the wh-phrase as an argument: they should notice that the filled-gap question has too many arguments. We hope that this work will allow us to probe more directly the structure of infants' syntactic representations of wh-questions at these two ages.

Although more direct evidence is needed, the results of the current study provide converging evidence for the hypothesis that infants rely on verb transitivity knowledge to identify filler-gap dependencies in sentences that contain them. But if this account is correct, we have a further puzzle: how can children learn verbs' distributional properties before they are able to identify instances of filler-gap dependencies, which obscure verb transitivity? That is, if children hear wh-object questions in such high frequency in their input, but do not yet recognize the structure of those clauses as underlyingly transitive, what prevents them from drawing faulty inferences about the syntactic and semantic properties of verbs in those sentences? A common proposal in the literature is that children need to “filter out” filler-gap dependencies and other types of non-basic clauses from the data they use for verb learning (Lidz & Gleitman, 2004a, 2004b, Pinker, 1984, 1989). In fact, it may be possible for learners to do so without identifying these clauses as non-basic: Perkins, Feldman & Lidz (2017) found that a Bayesian computational model could simultaneously infer the transitivity of verbs in child-directed speech and the parameters for filtering noisy or misleading data out of its input, without knowing why those data were misleading. Thus, a child might first ignore filler-gap dependencies in her input in order to arrive at stable perceptions of verb transitivity, and then to use verb transitivity to identify which sentences contain filler-gap dependencies.

Taken together with prior results, this work contributes to an incremental perspective on how children perceive and use the input to language acquisition. In order to identify the structure of sentences they hear, learners must rely on the immature linguistic knowledge they have acquired at their particular stage of development. We argue for a specific proposal implicating a tight relationship between verb argument structure and clause structure acquisition: namely, that argument structure knowledge acts as a catalyst for learners’ identification of filler-gap dependencies in sentences that contain them. As the current findings provide suggestive but not yet conclusive evidence in favor of this proposal, this work invites further investigation into the nature and development of filler-gap dependency representations in infancy.

References

Adani, F., Van der Lely, H. K., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, *120*(9), 2148–2166.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51*(1), 23–54.

Aoun, J., Hornstein, N., & Sportiche, D. (1981). Some aspects of wide scope quantification. *Journal of Linguistic Research*, *1*(3), 69–95.

Arnon, I. (2009). Rethinking child difficulty: The effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language*, *37*(01), 27–57.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.

Booth, A. E., & Waxman, S. R. (2009). A Horse of a Different Color: Specifying With Precision Infants' Mappings of Novel Nouns and Adjectives. *Child Development*, *80*(1), 15–22.

Cheng, L. L.-S. (2003). Wh-in-situ. *Glot International*, *7*(4), 103–109.

Chomsky, N. (1977). On Wh-Movement. In P. Culicover, A. Akmajian, & Wasow (Eds.), *Formal Syntax* (pp. 71–132). New York: Academic Press.

Cole, P., & Hermon, G. (1994). Is there LF wh-movement? *Linguistic Inquiry*, *25*(2), 239–262.

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

de Villiers, J., Flusberg, H. B. T., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *8*(5), 499–518.

Delle Luche, C., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the intermodal preferential looking paradigm: methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development*, *40*, 151–172.

Fenson, L., Bates, E., Dale, P. S., Hartung, J. P., Pethick, S. J., Reilly, J. S., … Thal, D. J. (1993). *The MacArthur communicative development inventories*. Singular Publishing Group.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 143–149.

Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, *27*(3), 339–374.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126.

Frazier, L., & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331–344.

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, *119*(1), 67–88.

Gagliardi, A., Mease, T. M., & Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15-and 20-month-olds. *Language Acquisition*, *23*(3), 1–27.

Goodluck, H., & Tavakolian, S. (1982). Competence and processing in children's grammar of relative clauses. *Cognition*, *11*(1), 1–27.

Hamburger, H., & Crain, S. (1982). Relative acquisition. *Language Development*, *1*, 245–274.

He, A. X., & Lidz, J. (2017). Verb learning in 14-and 18-month-old English-learning infants. *Language Learning and Development*, 1–22.

Hollich, G. (2005). *Supercoder: A program for coding preferential looking (Version 1.5).*

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Massachusetts Institute of Technology.

Jin, K., & Fisher, C. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. In *Proceedings of the 38th Boston University Conference on Language Development*. Boston, MA.

Kidd, E., & Bavin, E. L. (2002). English-speaking children’s comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, *31*(6), 599–617.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, *2*(1), 1–25.

Labelle, M. (1990). Predication, wh-movement, and the development of relative clauses. *Language Acquisition*, *1*(1), 95–119.

Lidz, J., & Gleitman, L. R. (2004a). Argument structure and the child’s contribution to language learning. *Trends in Cognitive Sciences*, *8*(4), 157–161.

Lidz, J., & Gleitman, L. R. (2004b). Yes, we still need universal grammar. *Cognition*, *94*(1), 85–93.

Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, *97*, 62–78.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*(4), 393–424.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: language input and acquisition*. Cambridge University Press.

Perkins, L., Feldman, N. H., & Lidz, J. (2017). Learning an input filter for argument structure acquisition. In *Proceedings of the 7th Workshop on Cognitive Modeling and Computational Linguistics*.

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

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.

Reinhart, T. (1998). Wh-in-situ in the framework of the Minimalist Program. *Natural Language Semantics*, *6*(1), 29–56.

Ross, J. R. (1967). *Constraints on variables in syntax.* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105–134.

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early Understanding of Subject and Object Wh-Questions. *Infancy*, *4*(3), 423–436.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, *13*(3), 272–281.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198.

Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition*, *4*(1–2), 5–48.

Suppes, P. (1974). The semantics of children's language. *American Psychologist, 29*(2), 103.

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163.

Tavakolian, S. (1981). The conjoined-clause analysis of relative clauses. *Language Acquisition and Linguistic Theory*, 167–187.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. G. De Villiers (Eds.), *Language Processing and Language Acquisition*. Dordrecht: Kluwer.

Wagers, M. W., & Phillips, C. (2014). Going the distance: memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304.

Appendix: List of Stimuli

Wh-questions (subject gap / object gap)

Which monkey is bumping the frog? / Which monkey is the frog bumping?

Which monkey is feeding the frog? / Which monkey is the frog feeding?

Which monkey is hugging the frog? / Which monkey is the frog hugging?

Which monkey is kissing the frog? / Which monkey is the frog kissing?

Which monkey is tickling the frog? / Which monkey is the frog tickling?

Which monkey is washing the frog? / Which monkey is the frog washing?

Relative Clauses (subject gap / object gap)

Find the monkey that’s bumping the frog. / Find the monkey that the frog is bumping.

Find the monkey that’s feeding the frog. / Find the monkey that the frog is feeding.

Find the monkey that’s hugging the frog. / Find the monkey that the frog is hugging.

Find the monkey that’s kissing the frog. / Find the monkey that the frog is kissing.

Find the monkey that’s tickling the frog. / Find the monkey that the frog is tickling.

Find the monkey that’s washing the frog. / Find the monkey that the frog is washing.