**Representations of non-local syntactic dependencies feed verb learning in infancy**

Laurel Perkins[1], Yuanfan Ying[2], Alexander Williams[3,4], and Jeffrey Lidz[3]

[1]Department of Linguistics, University of California Los Angeles
[2]Brain and Mind Institute, Chinese University of Hong Kong
[3]Department of Linguistics, University of Maryland
[4]Department of Philosophy, University of Maryland

**Conflict of Interest Statement**
The authors declare no potential conflict of interest.

**Data Availability Statement**
Upon acceptance for publication, the de-identified data that support the findings of this study will be made openly available in an OSF repository, and the URL and DOI will be provided.

**Ethics Approval Statement**
The authors certify that we have complied with APA ethical principles regarding research with human participants in the conduct of research presented in this manuscript. Research was conducted with approval from the University of Maryland and UCLA Institutional Review Boards.

**Funding Sources and Acknowledgments**
This work was supported by NSF grant #BCS-1551629 and NSF NRT award #DGE-1449815. We thank Gissell Alvarado, Mariam Aiyad, Sigríður Björnsdóttir, Angela Xiaoxue He, Mina Hirzel, Tyler Knowlton, Minqi Liu, Tara Mease, Taylor Sheely, the University of Maryland Project on Children's Language Learning, and the UCLA Language Lab for their assistance in preparing experimental materials and conducting these experiments.

**Research Highlights**

- We find that 19- to 21-month-olds represent *wh*-phrases (e.g., *what*) as non-local objects in *wh*-questions with unknown verbs (e.g., *What is the girl gonna gorp*?).
- We introduce a novel experimental paradigm, Violation of Fit, which measures how well an infant considers a particular sentence to fit with a particular scene.
- This test reveals that infants treat object *wh*-questions with a novel verb as transitive, and therefore a good fit to a causal event.
- This suggests that immediately after *wh*-dependency representations are first acquired, they are available to feed verb learning.

**Abstract**

The ability to represent both local and non-local syntactic dependencies emerges in an infant's second year of life, raising questions about how these early syntactic representations interact with language learning in other domains. Using *wh*-questions as our case study, we investigate how infants' syntactic dependency acquisition interacts with their early lexical development. Prior work finds that 18-month-olds represent fronted *wh*-phrases as non-local arguments in object *wh*-questions with known verbs. Here, we show that 19- to 21-month-olds (range: 18;29-21;26) do the same when interpreting unknown verbs. We introduce a novel Violation of Fit method, a cross-modal extension of the Violation of Expectations paradigm. Infants saw dialogues with novel verbs in object *wh*-questions (e.g. *What is the girl gonna gorp?*), transitive polar questions (*Is the girl gonna gorp the toy?*), or intransitive polar questions *(Is the girl gonna gorp?*). At test, infants viewed a causal event (e.g. a girl knocks over a tower) and we measured their attention as an indication of whether they considered the verbs to be a good fit for this type of event. Across the age range, we found that infants who heard *wh*-question dialogues attended similarly to the test events as infants who heard canonical transitive dialogues, and unlike infants who heard intransitive dialogues. Thus, 19- to 21-month-olds treat object *wh*-questions with a novel verb as transitive when relating them to scenes. This suggests that immediately after *wh*-dependency representations are first acquired, they are available to feed verb learning.

**Keywords:** language acquisition, syntax, non-local dependencies, *wh*-questions, verb learning, bootstrapping

## 1. Introduction

Children take their first steps towards acquiring a syntax for their native language before their second birthdays, with far-reaching consequences for their further linguistic development. Even rudimentary syntactic knowledge can scaffold the acquisition of sentence and word meaning (*syntactic bootstrapping:* Gleitman, 1990; Landau & Gleitman, 1985). One well-studied example is children's use of the basic predicate-argument relations in a sentence to bootstrap verb learning. By 20 months, infants who hear a verb in a transitive clause, with both a subject and an object, reliably infer that it labels a causal event, with an agent acting on a patient (see Fisher et al., 2019, for a review). Infants' ability to use subjects and objects in sentences to draw inferences about the event that a sentence describes, and thus narrow down the meanings of new verbs, likely plays an important role in vocabulary acquisition at this age.

But the use of syntax in early verb learning is made complicated by the variety of forms that predicate-argument relations can take. In English, the dependency between a verb and its object is canonically established in a local configuration, with the object (underlined) appearing immediately to the verb's right (1a). However, the same verb-object relation can also be established at a distance: in (1b), a fronted noun phrase (*what*) acts as the verb's object through a non-local *wh*-dependency.

(1) a. You fixed <u>your toy</u>.

b. <u>What</u> did you fix?

These dependencies can be expressed in different ways cross-linguistically, so learners must identify their form in the specific language that they are exposed to. Failure to do so may impact their inferences about the meanings of verbs in these sentences. Before infants can recognize the fronted object in a *wh*-question like (1b), they might misperceive this question as intransitive,

without an object. And before they know the meaning of the verb *fix,* this inaccurate sentence representation could lead to an inaccurate inference: they might think that *fixings* are not events that involve both agents and patients. Prior work argues that the prevalence of non-local and non-canonical predicate-argument dependencies in children's input could create sizable disruptions for verb and grammar learning in its early stages (Gleitman, 1990; Lidz & Gleitman, 2004; Perkins et al., 2022; Pinker, 1984).

Understanding the extent of this problem requires understanding when children represent non-local syntactic dependencies, and how robustly they can deploy this knowledge when it has been acquired. Prior work suggests that 18- to 20-month-olds represent the syntax of object *wh*-questions with known verbs (Gagliardi et al., 2016; Hirzel, 2022; Hirzel et al., 2020; Perkins & Lidz, 2021; Seidl et al., 2003). But this does not tell us whether this knowledge can immediately be deployed in bootstrapping the meanings of unknown verbs. If a 20-month-old does not already know the meaning or argument structure of the verb in (1b), will she still be able to recognize from the form of the sentence that this is an object *wh*-question, with *what* acting as a non-local object of *fix*? If so, then she might be able to use the transitivity of this question as evidence for what event in the world it describes, and thus what types of events *fixings* might be. But this rests on her ability to recognize that a non-local verb-object dependency is present, even without prior knowledge that the verb requires an object.

Through the case study of *wh*-dependencies, we investigate how infants' syntactic dependency acquisition interacts with their early lexical development. We find that by 19 to 21 months, infants can recognize these dependencies robustly enough to guide their interpretations of unknown verbs. Specifically, we find that infants across this age range treat novel verbs in object *wh*-questions as transitive when assessing their fit to causal events. This implies that *wh*-

question syntax is available to feed verb learning as soon as it is acquired, and indicates that children's earliest knowledge of non-local syntactic dependencies has immediate consequences for their further learning processes.

### *1.1 Local and Non-Local Dependencies in Verb Learning*

Children in their second year of life identify how predicate-argument relations are canonically expressed in their native language, and use those relations to draw inferences about verb and sentence meaning. English-learning 15- to 17-month-olds interpret transitive subjects in canonical pre-verbal positions as agents of events, and objects in canonical post-verbal positions as patients (Hirsh-Pasek & Golinkoff, 1996; Lidz et al., 2017; Perkins & Lidz, 2020). 19- to 24-month-olds who hear a transitive sentence with an unknown verb (e.g., *The boy is gorping the girl*) assign the verb a causal meaning, preferentially looking to a scene with an agent acting causally on a patient, such as a boy spinning a girl (Arunachalam et al., 2013; Arunachalam & Waxman, 2010; Messenger et al., 2015; Naigles, 1990; Yuan et al., 2012; Yuan & Fisher, 2009; a.o.). Similar behavior has been observed in 15-month-olds when tested in simplified designs (K.-S. Jin & Fisher, 2014).

It is unknown whether infants at these ages can draw these inferences when subjects and objects do not appear in their canonical positions local to the verb. Non-local predicate-argument dependencies are common in the speech that infants hear: *wh*-questions comprise approximately 15% of the input to English-learning 1-year-olds, and most have non-canonical word orders (Cameron-Faulkner et al., 2003; Stromswold, 1995). If children cannot reliably recognize these dependencies in sentences that contain them, then their immature representations of these sentences might provide misleading evidence about verb meaning. The frequency of *wh*-

dependencies and other non-canonical clause types in children's input therefore poses a challenge for bootstrapping as a reliable mechanism for early word learning (Gleitman, 1990; Lidz & Gleitman, 2004; Perkins et al., 2022; see also Pinker, 1984).

Previous findings suggest that infants identify non-local predicate-argument dependencies a few months after identifying clause arguments in their canonical positions. English-learning infants produce and robustly comprehend object *wh*-questions by the age of 20 months (Gagliardi et al., 2016; Rowland et al., 2003; Seidl et al., 2003; Stromswold, 1995), and might do so based on a correct representation of the non-local dependencies in these sentences (Gagliardi et al., 2016; Hirzel, 2022; Hirzel et al., 2020; Perkins & Lidz, 2021). To support this argument, Perkins & Lidz (2021) tested whether infants know that a fronted object *wh*-phrase cannot co-occur with a local post-verbal direct object, because they express the same grammatical relation: a sentence like **What did you fix your toy?* is ungrammatical because it contains too many objects. In a listening preference task, Perkins & Lidz (2021) found that 18-month-olds listened longer to basic declarative sentences in which transitive verbs took a required local object (*A dog! The cat should bump him* > **A dog! The cat should bump*), but listened longer to object *wh*-questions without a local object (*Which dog should the cat bump?* > **Which dog should the cat bump him?*). This suggests that they represented the *wh*-phrase as satisfying the verb's requirement for an object non-locally. Younger infants did not differentiate between these sentence types, suggesting that knowledge of the syntax of *wh*-dependencies emerges around 18 months, but not before.

If 18-month-olds represent non-local argument dependencies in *wh*-questions, then it is possible that a correct parse of these dependencies could drive their sentence and verb interpretation. The challenge that these dependencies pose for bootstrapping would then be

relatively short-lived, lasting from approximately 15 months, when infants first show evidence of learning verb meaning and argument structure, to approximately 18 months, when non-local dependencies are represented syntactically. Hirzel (2022) and Hirzel, Perkins, & Lidz (2020) suggest that this is the case, by showing that 19-month-olds incrementally interpret a fronted *wh*-phrase as the verb's direct object during online sentence processing. However, previous studies used *wh*-questions with verbs that infants already know. A crucial question is whether infants can recognize *wh*-dependencies with *unknown* verbs, so as to feed their inferences about what those verbs mean.

Listeners with mature knowledge of the system of syntactic dependencies in their language can parse *wh*-dependencies with entirely novel verbs. English-speaking adults who hear *What is she gonna gorp?* will parse this as an object *wh*-question by using the form of the sentence—the *wh*-word *what* and subject-auxiliary inversion—even without knowing the verb's meaning or argument-taking requirements. Is infants' emerging knowledge of *wh*-dependencies robust enough to support accurate parsing in this context? It certainly might not be. Possibly, when *wh*-dependencies are first acquired, infants rely on previously-learned verb knowledge to parse these dependencies online: knowing that a verb requires an object might help them parse a *wh*-phrase as that object, because this provides some signal for where the *wh*-dependency should be resolved. In this case, they might fail to parse these dependencies with verbs whose argument structure they do not know. The problem that non-local dependencies pose for bootstrapping could therefore persist past the point when the syntax of these dependencies is first acquired.

In summary, previous findings do not tell us when infants can parse the non-local dependencies in *wh*-questions reliably enough that the presence of these sentence types in their input can serve as useful input for verb learning, rather than a potentially disruptive source of

noise. On the one hand, 18-month-olds' ability to represent fronted *wh*-phrases as non-local arguments of known verbs makes it possible that a veridical parse of these dependencies might serve as the input to bootstrapping verb meanings after this age. On the other hand, the ability to identify non-local dependencies with known verbs does not necessarily imply that infants can parse these dependencies with verbs whose argument structure is currently unknown. If infants' recently-acquired *wh*-dependency knowledge is not robust enough to support reliable parsing with unknown verbs, then their mis-parses of these dependencies might continue to provide misleading evidence about verb meaning. Here, we provide a new empirical test to assess when children can parse *wh*-dependencies with novel verbs, in order to guide their inferences about what those verbs mean. This test reveals that 19- to 21-month-olds can indeed learn a novel verb's meaning on the basis of the transitivity of object *wh*-questions.

### *1.2 A Novel Empirical Paradigm: Violation of Fit*

The current work investigates when infants treat object *wh*-questions with unknown verbs, such (2a), as having the same number of arguments as sentences with canonical direct objects, such as (2b), while assessing their fit to scenes viewed as causal 2-participant events.

(2) a. What is the girl gonna *gorp*?

b. Is the girl gonna *gorp* the toy?

c. Is the girl gonna *gorp*?

If infants parse *what* as a non-local object of the verb in (2a) and therefore represent this sentence as transitive, then they should interpret the verb in a similar manner as infants who hear a transitive sentence like (2b), with the object in canonical position. In both cases, they should consider *gorping* to be a good label for a causal event. If, however, they mis-parse (2a) as

intransitive, then they should interpret the verb in a similar manner as infants who hear a sentence with no direct object, like (2c). In this case, *gorping* should not be a good label for certain types of causal events.

This design requires a measure of how infants view the goodness of fit between sentences and causal events. Prior tests of infants' verb learning have typically used preferential looking tasks that measure the degree to which infants who hear transitive vs. intransitive sentences prefer to look at events intended to be viewed as a causal, such as a boy spinning a girl, over events intended to be viewed as a non-causal, such as a boy and/or a girl moving independently (Arunachalam et al., 2013; Arunachalam & Waxman, 2010; K.-S. Jin & Fisher, 2014; Messenger et al., 2015; Naigles, 1990; Yuan et al., 2012; Yuan & Fisher, 2009). But a complication arises in these tasks: past results are inconclusive about whether infants think that a sentence with one argument is a bad fit for any particular stimulus scene. Infants who heard novel verbs in transitive sentences robustly preferred the intended causal events, but beyond an initial study by Naigles (1990), infants who heard intransitive sentences did not show reliable preferences in either direction, and instead were indifferent between the causal and non-causal scenes (Arunachalam et al., 2013; Arunachalam & Waxman, 2010; Noble et al., 2011; Yuan et al., 2012). There are several possible explanations for this indifference. Infants may have perceived the scenes under different representations than the experimenters intended: if they readily viewed a scene intended to be viewed as a boy spinning a girl under a different non-causal concept, such as two people playing, then both candidate scenes might be equally good fits for an intransitive clause (Arunachalam et al., 2016; Brandone et al., 2006; Pozzan et al., 2015). Furthermore, intransitive subjects can sometimes name patients of caused events: *The girl spun* can describe a girl's spinning, caused by a boy. If infants did not know whether to interpret the intransitive

subject as an agent or a patient, this would again make both scenes a plausible fit (Perkins, Knowlton, et al., 2025).

We address these issues in a novel experimental method. Because potential issues of indeterminacy in infants' scene percepts are made worse when two animate actors engage in ongoing actions, we present scenes in which a human agent performs a punctual action on an inanimate patient: e.g., a girl knocks over a tower of toys. Piloting with adults confirmed that these scenes are viewed in a way that makes them good fits for transitive descriptions. Furthermore, in the context of these scenes, an animate intransitive subject like that in (2c) unambiguously names the agent rather than the patient of the event. Within English and cross-linguistically, there is a robust tendency for clauses describing events viewed as changes to contain an argument naming the patient of that change (Williams, 2015; see also Fillmore, 1970; Levin & Hovav, 2005). These scenes might therefore lend themselves to intransitive descriptions where the sole argument names the patient (e.g., *Is the toy gonna gorp?*), but are unlikely to be viewed as good fits for intransitive descriptions of the sort in (2c), as confirmed by piloting with adults. Infants at 24-28 months, slightly older than the age tested here, show sensitivity to this distinction (Bunger & Lidz, 2004, 2008; Perkins, Mateu, et al., 2025; Scott & Fisher, 2009).

We present these scenes in a novel extension of the Violation of Expectations paradigm (Baillargeon et al., 1985; see Margoni et al., 2024 for a review), which we call ***Violation of Fit***. The logic underlying the Violation of Expectations technique is to familiarize infants with an event that can be encoded in a certain way, and then present a possible outcome that is either consistent or inconsistent with that encoding. If children attend differently to the inconsistent vs. consistent outcome, we infer that they had represented the event in the hypothesized manner. We extend the logic of this technique to a new cross-modal paradigm, in order to assess the

inferences that infants draw on the basis of familiarized sentences in the absence of concurrent events. During familiarization, we present dialogues containing a novel verb (e.g., *gorp*) in three between-subjects conditions, manipulating the syntactic frame: object *wh*-questions (2a), transitive polar questions (2b), or intransitive polar questions (2c). At test, we present a causal stimulus scene, and ask whether infants see *gorping*. Depending on infants' syntactic representations of the familiarization sentences, we predict that they will draw different inferences about what types of events *gorpings* might be. The test event will then be either a consistent or inconsistent fit to those partial meaning representations, resulting in different patterns of attention across familiarization conditions. If infants represent the familiarization sentences as transitive, then they should expect that *gorping* labels an event of an agent acting causally on a patient, and our stimulus event will fit well with those expectations. If infants represent the familiarization sentences as intransitive, then they might expect that *gorping* labels an action of an agent independent of a patient, and this causal event will not be a good fit.

This method is similar to preferential looking paradigms that investigate infants' bootstrapping inferences by first presenting a verb within a dialogue, and afterwards presenting a pair of candidate events (Arunachalam et al., 2013; Arunachalam & Waxman, 2010; Yuan & Fisher, 2009). But unlike past designs, our method does not give infants a choice between two events at test. Instead, we present a single test event, and ask whether or not infants perceive it as a good fit to *gorping*, given the sentences that they just heard. This taps more directly into our research question: it measures how well a particular scene and sentence fit together, rather than which of two scenes fits a sentence better.

Our dependent measure is looking time to the test video. There are two patterns that might be observed. Infants who think that the test event is inconsistent with *gorping* may respond

by becoming curious and investigating the scene, resulting in higher looking time compared to infants who perceive a consistent fit (e.g., Perez & Feigenson, 2022; Stahl & Feigenson, 2015; Wang & Baillargeon, 2008). Conversely, infants may respond to perceived inconsistency by disengaging, resulting in lower looking time compared to infants who perceive a consistent fit (e.g., K. Jin et al., 2024; Kuhlmeier et al., 2003). Like many other looking-time methods, we cannot predict in advance whether infants will look longer to a consistent vs. inconsistent fit, and the direction of this asymmetry might also change over development (Hunter & Ames, 1988; Kidd et al., 2012; Roder et al., 2000; Rose et al., 1982). What matters is whether infants' looking times differ as a function of their familiarization condition, and how those conditions pattern with respect to each other.

We predict that infants will readily view the test events as consistent with transitive polar questions of the form in (2b), and inconsistent with intransitive polar questions of the form in (2c). These conditions serve as controls to diagnose whether infants look longer to the consistent or inconsistent fit at the ages tested. We then compare the *wh*-question condition against these controls. If infants recognize that *what* is a fronted object in sentences like (2a), even with a verb that is unknown, then they will perceive these sentences as transitive and should view the test events as a consistent fit. This predicts similar patterns of attention for infants who hear *wh*-questions and those who hear transitive polar questions. However, if infants cannot recognize *what* as a fronted object when the verb is unknown, then they will perceive these sentences as intransitive and should view the test events as an inconsistent fit. This predicts that infants' behavior in the *wh*-question condition will more closely resemble the intransitive condition. Thus, by comparing infants' looking times to causal test events after hearing novel verbs in object *wh*-questions to their looking times after hearing those same verbs in transitive or

intransitive polar questions, we can determine whether infants take these *wh*-questions to be transitive when assessing their fit to scenes.

## 2. Method

### *2.1 Participants*

Participants included 72 typically-developing infants (36 males, 36 females) between the ages of 18;29-21;26 (mean age: 20;9). This age range was chosen to allow us to investigate development in the months after *wh*-question syntax is first acquired. Participants were recruited from two urban centers— the greater Washington, D.C. and Los Angeles areas— with the criterion that they heard English at least 80% of the time during their waking hours. An additional 25 infants were tested but excluded prior to analysis due to failure to complete the experiment (3), fussiness (6), parental interference (3), coder unable to see eyes (3), coder error (8), or equipment failure (1). Infants from the Washington, D.C. area were recruited online or over the phone through the University of Maryland's Infant and Child Studies Consortium database, with parental informed consent obtained according to the protocols of the University of Maryland's Institutional Review Board. Infants from the Los Angeles area were recruited online or over the phone through the University of Los Angeles Developmental Subject database, with parental informed consent obtained according to the protocols of UCLA's Institutional Review Board. Participants in the Washington, D.C. area were predominantly white. Participants in the Los Angeles area came from zip codes with predominantly white populations. Participants in both areas had predominantly college-educated parents.

Participants' total productive vocabulary was collected by parental report using the Words and Sentences MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al.,

**Figure 1.** Still images from test video stimuli: KNOCK-OVER (left), BREAK (right).

1993). Median total words produced were 91.5 (range: 5-657), median total verbs produced were 8 (range: 0-119), and median *wh*-words produced were 0 (range: 0-7).

### *2.2 Materials*

Familiarization stimuli consisted of videos of 11-second dialogues between native English speakers. Each dialogue contained a novel verb in four sentences of the same type: transitive polar questions, intransitive polar questions, or object *wh*-questions (see the Appendix for full text). Two sets of dialogues were filmed for each sentence type, containing different novel verbs (*blick* and *gorp*).

Test visual stimuli consisted of videos of a woman performing punctual actions with inanimate objects: knocking over a tower, or breaking a breadstick in half (Figure 1). Three tokens of each event were filmed and edited to be 5 seconds long. Each test trial contained six repetitions of a single event type, with tokens presented in pseudo-random order. Test audio stimuli were recorded by a female native English speaker in a child-directed register, and were combined with videos to create the trial structure in Table 1. Within each sentence-type condition, two sets of test trials were created, crossing novel verb (*blick* or *gorp*) with event type (KNOCK-OVER or BREAK). A silent video of rotating shapes was prepared as an inter-trial attention-getter stimulus, and an 87-second video of moving toys set to music was prepared as an attention-getter halfway through the experiment.

| Phase | Sample audio, *wh*-question condition | Video | |
|---|---|---|---|
| **Familiarization** (4 trials x 11 s) | - What is the girl gonna gorp?<br>- Yeah, what is she gonna gorp?<br>- Right, what is she gonna gorp?<br>- No idea! What is she gonna gorp? |  | 2 actors conversing |
| **Test** (4 trials x 35 s) | What is she gonna gorp? | | 3-s black screen |
| | Ooh, look! … Did you see gorping?<br>Hey, wow! … Did you see gorping? |  | Girl knocks over tower (6 times on loop) |

**Table 1.** Experimental structure, for one sample verb-event pair. Each child was exposed to the procedure twice, once for each event type paired with a different novel verb. Sample audio is for the *wh*-question condition; the other groups saw identical video, but heard transitive or intransitive polar questions instead of *wh*-questions.

To norm the test videos, we solicited spontaneous descriptions from 15 adult university students recruited from the University of Maryland's SONA system database, with informed consent obtained in accordance with the University of Maryland's Institutional Review Board. The task took approximately 5 minutes, and participants received course credit for participating. Participants viewed one token of each test event, and were instructed to write down three sentence-length descriptions of what happened in the video. Of the responses that mentioned the main action in the scene, 96% (65 out of 68) described the action using transitive syntax, and all participants provided at least one transitive description in the form of (2b). No participants produced an intransitive description in the form of (2c), with the subject naming the agent of the action; the few intransitive descriptions took a different form, with the subject naming the patient (e.g., *The tower falls over*). We therefore gain confidence in our assessment that people readily view these stimuli in a way that makes them a good fit for the transitive sentences and a poor fit for the intransitive sentences in our experiment.

### *2.3 Procedure*

Infants sat in front of a widescreen television in a dimly-lit room, either on their parents' lap or in a high chair with their parent in a chair next to them. Conditions were carefully matched across

the two testing locations. At the University of Maryland, stimuli were presented on a 51-inch television mounted below a video camera, with infants seated 66 inches away. At UCLA, stimuli were presented on a 46-inch widescreen television mounted above a video camera, with infants seated 42 inches away to match their viewing angle as closely as possible to that at Maryland. Parents were instructed to close their eyes and refrain from speaking to their child or directing their child's attention.

Stimuli were presented using the Habit software (Cohen et al., 2004). An experimenter in an adjacent room watched a live feed from the video camera and coded infants' gaze by pressing a key when the infant was looking at the screen, and releasing it when the infant looked away. Pan and zoom of the camera were controlled remotely to keep the infant's face in the video frame. Videos were recorded and saved for each experimental session.

The experiment consisted of two portions, each structured according to Table 1. First, the silent attention-getter stimulus was displayed. Once the infant fixated the attention-getter, the experimenter began the familiarization phase. This consisted of four trials with a fixed length of 11 seconds. All four trials presented a dialogue with the same novel verb, with sentence type (*wh*-question, transitive, or intransitive) differing across conditions. Infants were randomly assigned to one of the three sentence-type conditions. At the end of each trial, the attention-getter stimulus was displayed, and the next trial began when the infant re-oriented towards the television.

After the four familiarization trials, the test phase began. This also consisted of four trials, all presenting tokens of the same test event (either KNOCK-OVER or BREAK). Each trial began with a 3-second blank screen presented alongside audio containing the familiarized novel verb, using the familiarized sentence type. Up to six repetitions of the test event then played on loop

alongside audio of the same novel verb in neutral syntax; this audio did not differ by condition. Infants' looking times were only coded while the test event was present on the screen, which ensured that test looking times were measured during the portion of the trial where infants in each condition viewed exactly the same audiovisual stimulus. Unlike the familiarization trials, test trials were infant-controlled: they ended if either the maximum trial length (35 seconds) was reached, or if the infant looked away from the television for 2 seconds. The attention-getter stimulus was again displayed between trials, and a new trial began when an infant re-oriented towards the television.

Both events were tested within-participants, paired with different novel verbs (*blick* or *gorp*). Each infant was first exposed to the familiarization and test phases with one verb and event type, and then repeated the procedure with the second verb and event type. Verb-event pairing and order were counterbalanced across participants. Each infant heard the same familiarization sentence type for both portions of the experiment. Between the two portions, the 87-second video of moving toys was displayed to refocus attention.

**3. Results**

To confirm online coding reliability, videos of seven experimental sessions were recoded frame-by-frame by a different experimenter. Intercoder agreement was above 86% for all videos.

To investigate the developmental trajectory of infants' looking behavior across the age range that we tested, we conducted linear mixed-effects regressions using the *lme4* package in R (Bates et al., 2015). Our dependent measure was log-transformed looking times during each trial.[1] Factor contrasts were sum-coded and significance testing was performed through likelihood ratio tests, comparing a model containing the effect against a model that differed only

in that it lacked the relevant effect.

We first analyzed infants' familiarization looking times to confirm that infants across conditions displayed no significant looking differences during the familiarization phase. We fit a mixed-effects regression model with fixed effects of age in days and condition (*wh*-question, transitive, or intransitive), as well as their interaction. The maximal random effects structure that converged included a random intercept for participant. We found no significant main effects of condition ($\chi^2(2) = 0.38$, $p = 0.83$) or age ($\chi^2(1) = 0.72$, $p = 0.40$), and no interaction of condition and age ($\chi^2(2) = 3.93$, $p = 0.14$). Thus, infants throughout the age range attended similarly to the familiarization dialogues in each condition.

We then analyzed infants' looking times during the test trials to determine whether their behavior varied by condition when viewing our causal test events. We again fit a mixed effects regression model with fixed effects of age, condition, and their interaction. The maximal random effects structure that converged included random intercepts for participant and trial. We found no significant main effects of condition ($\chi^2(2) = 2.11$, $p = 0.35$) or age ($\chi^2(1) = 0.24$, $p = 0.62$), but a significant interaction of condition and age ($\chi^2(2) = 8.72$, $p < 0.05$). Posthoc analyses on the slopes of the age trends for each condition revealed a significant difference between the transitive and intransitive conditions ($t(71.1) = 2.48$, $p < 0.05$, Satterthwaite's approximation for degrees of freedom): across the age range, infants registered the asymmetry in goodness of fit between our causal test events and the transitive vs. intransitive polar questions. Moreover, we found that the *wh*-question condition differed significantly from the intransitive condition ($t(71.1) = 2.78$, $p < 0.05$), but not from the transitive condition ($t(71.1) = 0.67$, $p = 0.78$). To visualize this interaction, we plot average raw looking times by condition for each infant as a function of age in Figure 2, along with simple linear regression trend lines. Younger infants

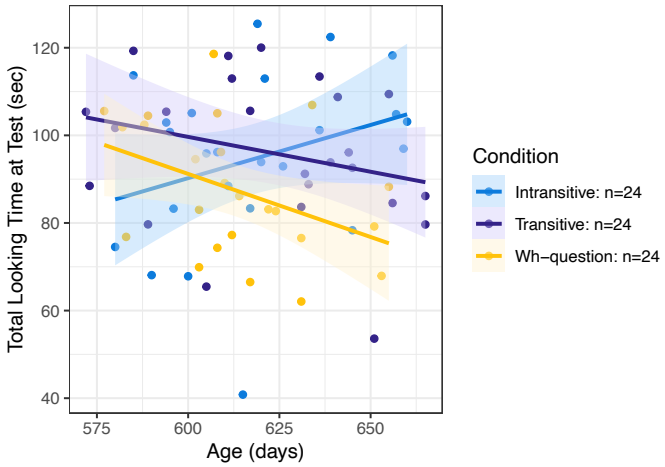


**Figure 2.** Average total looking time during the test phase, by condition and age. Points represent average looking times across all test trials for each individual infant. Lines display simple linear regression estimates.

attended longer to the test events after hearing *wh*-question and transitive dialogues, compared to intransitives; older infants attended longer after hearing intransitive dialogues compared to the others. Crucially, despite this change in the absolute direction of infants' looking preferences over the age range, we found that their relative patterns of attention differed consistently across the conditions: their behavior in the *wh*-question condition patterned like their behavior in the transitive condition, and unlike their behavior in the intransitive condition.

We further investigated whether event type (KNOCK-OVER or BREAK) or lab site (UCLA or Maryland) affected infants' behavior by fitting additional mixed-effects regression models which added these factors as fixed effects, as well as their interactions with condition and age. Random effects included a random intercept for participant and a random intercept for trial. The interaction of condition and age was again significant in the model with lab site as a predictor ($\chi^2(2) = 10.10$, $p < 0.01$). This model also revealed a main effect of lab ($\chi^2(1) = 4.84$, $p < 0.05$), indicating that infants looked overall longer at the UCLA site than at the Maryland site. The model with event type as a predictor again revealed a significant interaction of condition and age ($\chi^2(2) = 8.72$, $p < 0.05$), as well as a main effect of event type ($\chi^2(1) = 4.87$, $p < 0.05$), indicating

that infants looked overall longer to the KNOCK-OVER video than to the BREAK video. No other significant main effects or interactions were found in either model (all *p*s > 0.11). Importantly, there were no interactions of either event type or lab site with condition or age. Thus, the asymmetries in infants' looking times across conditions were exhibited to the same extent at both sites and for both event types.

## 4. Discussion

Prior work argues that the ability to represent non-local syntactic dependencies in *wh*-questions emerges in the second year of life, raising questions about how these early syntactic representations interact with language learning in other domains (Gagliardi et al., 2016; Hirzel, 2022; Hirzel et al., 2020; Perkins & Lidz, 2020, 2021; Seidl et al., 2003). We investigate the consequences of this knowledge for word learning in the months immediately after *wh*-dependency representations first emerge. We find that 19- to 21-month-olds recognize a fronted *wh*-phrase as a non-local object of a verb even when the verb is unknown, and accurately use these representations to guide their inferences about what the verb might mean. In a novel experimental paradigm, infants who were familiarized to new verbs in object *wh*-questions showed similar patterns of attention as infants who were familiarized to the same verbs in canonical transitive sentences, when they were later asked whether a causal event provided a referent for those verbs. Both groups attended differently from infants who were familiarized with intransitive sentences. This suggests that infants represented the non-local verb-object dependency in the *wh*-questions: they treated object *wh*-questions with novel verbs as transitive rather than intransitive, and therefore a good fit for causal events.

We also found that infants' patterns of attention differed by age. Younger infants who

heard intransitive familiarization looked less long at test than infants who heard transitive or *wh*-question familiarization. Older infants who heard intransitive familiarization looked longer at test than infants in the other conditions. That is, younger infants appeared to respond to the inconsistent fit between intransitive syntax and our causal test events by disengaging, whereas older infants responded to this perceived inconsistency by investigating the scene longer. This pattern suggests differences in task difficulty for infants at different points across our age range (Kidd et al., 2012). Crucially, what matters in this design is not the absolute direction of infants' looking asymmetries, but whether their relative preferences are the same or different across conditions at a given age. We find that infants across the age range differentiated object *wh*-questions from intransitives, and treated them similarly to canonical transitives when assessing their fit to scenes.

These findings contribute to the wider literature on infants' use of syntax in verb learning, with important implications for the reliability of syntactic bootstrapping early in development. Because it is not trivial for learners to identify the forms that non-local syntactic dependencies can take in their language, prior work has argued that the prevalence of these sentence types in children's input introduces a potentially serious source of noise into the body of evidence that they have available for bootstrapping verb meanings (Gleitman, 1990; Lidz & Gleitman, 2004; Perkins et al., 2022). If infants fail to parse object *wh*-questions as transitive when they contain unknown verbs, then their incomplete representations of these sentences might provide misleading information about the meanings of verbs in these sentences. The current results suggest that this problem is resolved early in development. In the months immediately after infants show the ability to represent fronted *wh*-phrases as objects of known verbs, we find that that they can also deploy this knowledge to recognize these dependencies with new verbs, and

draw appropriate inferences about what events these verbs might label. This suggests that a correct parse of *wh*-dependencies is available for bootstrapping verb meanings by 19 months, only a few months after infants first begin learning verb meaning and argument structure (Hirsh-Pasek & Golinkoff, 1996; K.-S. Jin & Fisher, 2014; Lidz et al., 2017; Perkins & Lidz, 2020). Further work is needed to probe how infants' bootstrapping is affected by the presence of *wh*-questions in their input prior to this age, a topic that has been studied computationally (Perkins et al., 2022).

A secondary contribution of the current study is to show that 19- to 21-month-olds consider certain intransitive clauses to be a poor fit for scenes viewed as causal events. This has not been found reliably in previous preferential looking tasks, where infants who heard intransitive sentences have been ambivalent between intended causal and non-causal scenes (Arunachalam et al., 2013; Arunachalam & Waxman, 2010; Noble et al., 2011; Yuan et al., 2012). Past inconclusive results may have derived from ambiguity in how infants represented the stimulus scenes, which the current work attempted to control for (Arunachalam et al., 2016; Brandone et al., 2006; Perkins, Knowlton, et al., 2025; Pozzan et al., 2015). But infants' previous indifference also raised the possibility that children at this age do not use intransitive syntax to constrain their inferences about what events a new verb might label. Perhaps when hearing *The girl gorped*, children believe that *gorpings* might label any event in which a girl is a participant, whether or not other participants are also present (Fisher et al., 2019; Williams, 2015). Our results argue against this account: the infants in our study registered an inconsistency between an intransitive clause with *the girl* as subject and a scene viewed as a causal 2-participant event, in which a girl effects a change to a patient. This inconsistency reflects a robust cross-linguistic tendency for clauses describing events viewed as changes to realize the patient of that change. It

is unlikely for single events of change to be described by a clause whose sole argument names, not the patient, but the agent of the change: for instance, **The fire melted* cannot express the meaning "the fire melted something" (Fillmore, 1970; Levin & Hovav, 2005; Williams, 2015). Our results suggest that infant behavior respects this distinction, even at earlier ages than shown in prior work (Bunger & Lidz, 2004, 2008; Perkins, Mateu, et al., 2025; Scott & Fisher, 2009). We find that 19- to 21-month-olds can draw fine-grained inferences about the meanings of verbs in clauses with a single argument, in a way that is consistent with patterns of argument realization observed across diverse languages. These findings therefore contribute an early data point on infants' bootstrapping abilities from intransitive clauses, and invite further work investigating learners' initial expectations about the correspondences between intransitive syntax and meaning at the onset of verb learning.

One reason that we find an effect with intransitives at this age, where previous studies did not, may lie in our novel method. We introduce and validate the Violation of Fit paradigm, which extends the Violation of Expectations method cross-modally: we measure children's predictions about upcoming events solely on the basis of their syntactic representations of sentences. This provides a potentially more sensitive alternative for younger populations than the Preferential Looking method commonly used in earlier bootstrapping studies. An important innovation is that the Violation of Fit paradigm does not require children to compare two scenes at test, which both simplifies the child's task and provides a more direct probe into our research question. Whereas the Preferential Looking method can only measure whether children consider one scene to be a better fit than another for a particular sentence, the Violation of Fit method measures how well individual scenes and sentences fit together. We believe that this provides a useful new method for future research to test hypotheses about the ranges of interpretations that children have

available for sentences that they hear.

In summary, these results provide new evidence for the interaction between early syntactic dependency acquisition and word learning in infancy. We find that 19- to 21-month-olds not only represent non-local syntactic dependencies in *wh*-questions with known verbs, but can also use correct representations of these dependencies to interpret unknown verbs. This suggests that immediately after *wh*-question syntax is acquired, children can deploy this knowledge to guide their inferences about word meaning. And this has broader implications for the interconnectedness of syntactic representations in development: even at some of the earliest stages of syntax acquisition, children's developing syntactic knowledge feeds their learning in other linguistic domains.

**References**

Arunachalam, S., Escovar, E., Hansen, M. A., & Waxman, S. R. (2013). Out of sight, but not out of mind: 21-month-olds use syntactic information to learn verbs even in the absence of a corresponding event. *Language and Cognitive Processes*, *28*(4), 417–425.

Arunachalam, S., Syrett, K., & Chen, Y. (2016). Lexical disambiguation in verb learning: Evidence from the conjoined-subject intransitive frame in English and Mandarin Chinese. *Frontiers in Psychology*, *7*, 138.

Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*(3), 442–446.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208.

Bates, D., Maechler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Brandone, A., Addy, D. A., Pulverman, R., Golinkoff, R. M., & Hirsh-Pasek, K. (2006). One-for-one and two-for-two: Anticipating parallel structure between events and language. *Proceedings of the 30th Annual Boston University Conference on Language Development*, 36–47.

Bunger, A., & Lidz, J. (2004). Syntactic bootstrapping and the internal structure of causative events. *Proceedings of the 28th Annual Boston University Conference on Language Development*, 74-85.

Bunger, A., & Lidz, J. (2008). Thematic relations as a cue to verb class: 2-year-olds distinguish unaccusatives from unergatives. *University of Pennsylvania Working Papers in Linguistics*, *14*(1), 43-55.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). *Austin: University of Texas*.

Fillmore, C. J. (1970). The grammar of "hitting" and "breaking." In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English Transformational Grammar* (pp. 120–133). Ginn and Company.

Fisher, C., Jin, K.-S., & Scott, R. M. (2019). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77.

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.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105–124). The MIT Press.

Hirzel, M. R. (2022). *Island Constraints: What is there for children to learn?* PhD Thesis, University of Maryland, College Park.

Hirzel, M. R., Perkins, L., & Lidz, J. (2020). *19 Month-Olds Represent and Incrementally Parse Filler-Gap Dependencies* [Poster]. 3rd Annual CUNY Human Sentence Processing Conference, Amherst/Online. osf.io/v3k27

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*.

Jin, K., Ting, F., He, Z., & Baillargeon, R. (2024). Infants expect some degree of positive and negative reciprocity between strangers. *Nature Communications*, *15*(1), 7742.

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

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One*, *7*(5), e36399.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402–408.

Landau, B., & Gleitman, L. R. (1985). *Language and Experience: Evidence from the Blind Child*. Harvard University Press.

Levin, B., & Hovav, M. R. (2005). *Argument realization*. Cambridge University Press.

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

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

Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review*, *131*(3), 716.

Messenger, K., Yuan, S., & Fisher, C. (2015). Learning verb syntax via listening: New evidence from 22-month-olds. *Language Learning and Development*, *11*(4), 356–368.

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(2), 357–374.

Noble, C. H., Rowland, C. F., & Pine, J. M. (2011). Comprehension of Argument Structure and Semantic Roles: Evidence from English-Learning Children and the Forced-Choice Pointing Paradigm. *Cognitive Science*, *35*(5), 963–982.

Perez, J., & Feigenson, L. (2022). Violations of expectation trigger infants to search for explanations. *Cognition*, *218*, 104942.

Perkins, L., Feldman, N. H., & Lidz, J. (2022). The power of ignoring: Filtering input for argument structure acquisition. *Cognitive Science*, *46*(1), e13080.

Perkins, L., Knowlton, T., Williams, A., & Lidz, J. (2025). Thematic Content, Not Number Matching, Drives Syntactic Bootstrapping. *Language Learning and Development*, *21*(2), 142–172.

Perkins, L., & Lidz, J. (2020). Filler-gap dependency comprehension at 15 months: The role of vocabulary. *Language Acquisition*, *27*(1), 98–115.

Perkins, L., & Lidz, J. (2021). 18-month-old infants represent non-local syntactic dependencies. *Proceedings of the National Academy of Sciences*, *118*(41), e2026469118.

Perkins, L., Mateu, V., & Hyams, N. (2025). 28-Month-Olds Use Inferred Thematic Relations to Bootstrap Intransitive Verb Meanings. *Proceedings of the 49th Annual Boston University Conference on Language Development*, 561–574.

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

Pozzan, L., Gleitman, L. R., & Trueswell, J. C. (2015). Semantic ambiguity and syntactic bootstrapping: The case of conjoined-subject intransitive sentences. *Language Learning and Development*, *12*(1), 14-41.

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4), 491–507.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*(5), 704.

Rowland, C. F., Pine, J. M., Lieven, E. V., & Theakston, A. L. (2003). Determinants of acquisition order in wh-questions: Re-evaluating the role of caregiver speech. *Journal of Child Language*, *30*(3), 609–635.

Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, *24*(6), 777–803.

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

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

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

Wang, S., & Baillargeon, R. (2008). Can infants be "taught" to attend to a new physical variable in an event category? The case of height in covering events. *Cognitive Psychology*, *56*(4), 284–326.

Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge University Press.

Yuan, S., & Fisher, C. (2009). “Really? She Blicked the Baby?”: Two-Year-Olds Learn Combinatorial Facts About Verbs by Listening. *Psychological Science*, *20*(5), 619–626.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*(4), 1382–1399.

## Appendix

| | Condition | | |
|---|---|---|---|
| | ***Wh*-question** | **Transitive polar question** | **Intransitive polar question** |
| 1 | - Hey, what is the girl gonna gorp?<br>- Hmm, what is the girl gonna gorp?<br>- Yeah, what is she gonna gorp?<br>- I don't know! What is she gonna gorp? | - Hey, is the girl gonna gorp the toy?<br>- Hmm, is the girl gorp the toy?<br>- Yeah, is she gonna gorp it?<br>- I don't know! Is she gonna gorp the toy? | - Hey, is the girl gonna gorp?<br>- Hmm, is the girl gorp?<br>- Yeah, is she gonna gorp?<br>- I don't know! Is she gonna gorp? |
| 2 | - Did you hear? What is the girl gonna gorp?<br>- Huh, what is she gonna gorp?<br>- Uh-huh! What is she gonna gorp?<br>- Beats me. What is she gonna gorp? | - Did you hear? Is the girl gonna gorp the toy?<br>- Huh, is she gonna gorp it?<br>- Uh-huh! Is she gonna gorp the toy?<br>- Beats me. Is she gonna gorp it? | - Did you hear? Is the girl gonna gorp?<br>- Huh, is she gonna gorp?<br>- Uh-huh! Is she gonna gorp?<br>- Beats me. Is she gonna gorp? |
| 3 | - Ooh, what is the girl gonna gorp?<br>- Yeah, what is she gonna gorp?<br>- Right, what is she gonna gorp?<br>- No idea! What is she gonna gorp? | - Ooh, is the girl gonna gorp the toy?<br>- Yeah, is she gonna gorp the toy?<br>- Right, is she gonna gorp it?<br>- No idea! Is she gonna gorp it? | - Ooh, is the girl gonna gorp?<br>- Yeah, is she gonna gorp?<br>- Right, is she gonna gorp?<br>- No idea! Is she gonna gorp? |
| 4 | - Are you ready? What is the girl gonna gorp?<br>- Whoa, what is the girl gonna gorp?<br>- Yeah, what is she gonna gorp?<br>- Wow, let's see! What is she gonna gorp? | - Are you ready? Is the girl gonna gorp the toy?<br>- Whoa, is the girl gonna gorp the toy?<br>- Yeah, is she gonna gorp it?<br>- Wow, let's see! Is she gonna gorp the toy? | - Are you ready? Is the girl gonna gorp?<br>- Whoa, is the girl gonna gorp?<br>- Yeah, is she gonna gorp?<br>- Wow, let's see! Is she gonna gorp? |

**Table A1.** Full experimental dialogues for the four familiarization trials in each condition, for one of two novel verbs. The same text was used for the second novel verb (*blick*).

[1] The main findings from these linear mixed-effects regressions remain the same if conducted on raw rather than log-transformed looking times.