ARTICLE IN PRESS

Cognition xxx (xxxx) xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

The hunt for structure-dependent interpretation: The case of Principle C☆

Jeffrey Lidz *, Cynthia Lukyanenko, Megan Sutton

Department of Linguistics, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742, USA

ARTICLE INFO

Keywords:
Binding
Principle C
Syntactic processing
Processing speed
Structure dependent interpretation

ABSTRACT

Jacques Mehler's earliest work concerned the independence of syntactic and semantic representations in adult sentence understanding, probing for independent contributions of sentence structure and sentence meaning in the psychological processes that underlie linguistic perception (e.g., Mehler, 1963; Mehler & Miller, 1964). The bulk of his career was spent pioneering the study of infants' linguistic cognition. In this paper, we bring these two streams together, using data from a suite of infant looking tasks to probe the syntactic representations that underlie sentence understanding for 30-month-olds. Each participant completed a battery of 3 tasks: one measuring knowledge of Principle C, one measuring lexical access speed and one measuring syntactic processing. We find that variability in performance on a Principle C task is predicted by variability in vocabulary, but not by either lexical access speed or a new measure of syntactic integration. Successful deployment of Principle C in 30-month-olds may therefore depend on factors related to vocabulary, but distinct from either lexical access or structure building. Identification of such factors remains an important goal for future work.

1. Introduction

Child language learners have routinely been shown to comprehend far more linguistic information than they produce (Christophe & Morton, 1998; Goodman & Jusczyk, 2000; Smith, 1975; Gerken, Landau, & Remez, 1990; Lidz, Waxman, & Freedman, 2003; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Seidl, Hollich, & Jusczyk, 2003; and many others). Given that comprehension precedes production, a major effort in the study of language acquisition is to find evidence of children's knowledge of linguistic structure that they do not yet exhibit in their own productions. A significant volume of research has thus endeavored to probe children's complex linguistic knowledge at increasingly young ages. Methodologies employed for these types of investigation are designed to alleviate extra-linguistic demands, in order to more accurately diagnose linguistic knowledge (Golinkoff et al., 1987; Kemler Nelson et al., 1995; Werker, Cohen, Lloyd, Casasola, & Stager, 1998).

Because work with the youngest learners often depends on looking time measures, studies on the acquisition of syntax often rely on measures that probe aspects of meaning that depend on syntax (De Carvalho, He, Lidz, & Christophe, 2019; Lidz et al., 2003; Lukyanenko, Conroy, & Lidz, 2014; Seidl et al., 2003; Shi, Legrand, & Brandenberger, 2020; Yuan & Fisher, 2009). Such work runs a risk, however, of not probing syntactic knowledge directly. Children might arrive at the correct understanding of an utterance without relying on the same syntactic representations as adults (Perkins & Lidz, 2020). And because one of the primary functions of syntax is to carry meaning, there are few methods that allow us to identify syntactic representations that do not rely on interpretation.

In adult psycholinguistics, the hunt for effects of syntactic representations that were independent of meaning began at the dawn of the modern study of psycholinguistics. In a series of papers, Mehler and colleagues attempted to dissociate the effects of syntax from the effects of semantics (Mehler, 1963; Mehler & Carey, 1967; Mehler & Carey, 1968; Mehler & Miller, 1964), probing for effects of silent aspects of syntactic structure. While some of the conclusions of this work were later criticized (Fodor, Bever, & Garrett, 1967; Townsend & Bever, 2001), the study of how syntactic form interacts with other aspects of mind that contribute to sentence understanding remains an important area of study in both adults and children (Momma & Phillips, 2018; Omaki & Lidz, 2015; Trueswell, Sekerina, Hill, & Logrip, 1999; Wagers, 2014). When it comes to the youngest learners, however, there has been little attempt to dissociate syntactic processing from semantic interpretation.

We present here a probe for syntactic mechanisms underlying interpretation in the domain of "Principle C" effects at 30 months.

☆ This paper is a part of special issue "Special Issue in Honour of Jacques Mehler, Cognition's founding editor".

* Corresponding author.
E-mail address: jlidz@umd.edu (J. Lidz).

https://doi.org/10.1016/j.cognition.2021.104676
Received 25 August 2020; Received in revised form 10 March 2021; Accepted 12 March 2021

Principle C is part of a theory explaining constraints on referential dependencies by virtue of the syntactic relations between Noun Phrases. Roughly, it characterizes speakers' inability to create a referential dependency between a pronoun and a name in its scope. Thus, in sentences like (1), the pronoun *she* and the name *Katie* cannot co-refer.

(1) She is patting Katie

This paper has two goals, one empirical and one methodological. Empirically, we aim to diagnose whether children's apparent sensitivity to constraints on reference derives from the same structural source as adults' sensitivity. Methodologically, we develop a measure of syntactic structure building, and aim to understand whether correlations between measures of understanding in different syntactic constructions allow us to infer shared mechanisms for representing those constructions. Utilizing the preferential looking paradigm, we analyze individual variation with respect to four aspects of children's linguistic abilities – their vocabulary, the speed of lexical access, their syntactic structure building and their interpretation of pronouns in Principle C contexts. Our reasoning is that if two abilities draw on the same linguistic resources, then we might see correlations between performance on tasks probing those abilities. Of course, a failure to find such correlations is not necessarily evidence that the two tasks do not rely on the same resources, just that the variability we see across tasks is not attributable to those resources. We find that variability in performance on a Principle C task is predicted by variability in vocabulary, but not by either lexical access speed or our new measure of syntactic integration, leaving open the question of how vocabulary relates to success with Principle C.

2. Background

Principle C is part of Binding Theory, a set of three structure-dependent constraints on the interpretive relations between noun phrases (NPs) in a sentence. These constraints are defined in (2) as restrictions on where anaphoric dependencies between two NPs can and cannot occur. Necessary terminology of *binding* and *c-command* are defined in (3 and 4), respectively. Principle A governs the use of anaphors (reflexive pronouns and reciprocals), Principle B governs the use of pronominals (non-reflexive pronouns), and Principle C (2c) governs the use of R-expressions, i.e. those NPs that are not subject to Principles A or B (e.g. *the cat*, *cookies*, *Jennifer, every student*). The constraints of Binding Theory represent concise descriptive generalizations covering a wide range of facts about the environments in which grammatically enforced referential dependencies are and are not possible (e.g. Chomsky, 1981; Lakoff, 1968; Langacker, 1966; Lasnik, 1976; Ross, 1967).

(2) a. Principle A: an anaphor must be bound in its governing category.
b. Principle B: a pronoun must be free in its governing category.
c. Principle C: an R-expression must be free.

(3) A node α binds a node β iff:
a. α and β are co-indexed,
b. α c-commands β.

(4) A node α c-commands a node β iff
a. neither node dominates the other
b. the first branching node dominating α dominates β.

A node's index is a pointer to its referent. If two nodes are coindexed, coreference between them is obligatory.

As stated in (2c), Principle C restricts the set of possible interpretations available for sentences containing an R-expression. Specifically, it blocks a bound interpretation when an R-expression occurs within the c-command domain of another NP. In (5a-c), the NP *Katie* falls outside the c-command domain of the pronoun *she*. Consequently, coindexation between these two NPs does not yield a binding relation and so Principle C is satisfied. In (5d), however, the NP *Katie* does occur within the c-command domain of the pronoun *she*, and so if these NPs are coindexed, Principle C is violated. Thus, *she* and *Katie* must be interpreted as disjoint in reference.

(5) a. While Katie_1 was in the kitchen, $\text{she}_{1/2}$ baked cookies.
b. While $\text{she}_{1/2}$ was in the kitchen, Katie_2 baked cookies.
c. Katie_1 baked cookies while $\text{she}_{1/2}$ was in the kitchen.
d. $\text{She}_{1/*2}$ baked cookies while Katie_2 was in the kitchen.

In sentences like (5a-c), where Principle C is satisfied, a host of discourse related factors contribute to the likelihood of coreference between the pronoun and the name. Such factors are essentially mooted in cases where Principle C does apply, however. For example, changing the verb impacts the likely interpretation of (6a), with *bought* encouraging disjoint reference and *tasted* encouraging coreference, but not (6b):

(6) a. While Katie_1 was in the kitchen, $\text{she}_{1/2}$ delivered/bought/tasted the cookies.
b. $\text{She}_{1/*2}$ delivered/bought/tasted the cookies while Katie_2 was in the kitchen.

The specification of c-command as the correct syntactic relation for defining binding is discussed at length by Reinhart (1976, 1983)[1] (see also Chomsky, 1981; Chomsky, 1986; Fiengo, 1977). Assuming the correctness of this analysis, adult-like interpretation of possible binding relations between NPs inherently requires a representation of the hierarchical structure of sentences.

From the perspective of language acquisition, then, correct application of Principle C will depend on several factors. First, children must be able to determine which Noun Phrases fall into each of the three categories – anaphor, pronominal or R-expression. This determination will require the child to have sufficient linguistic experience to categorize words appropriately (see Orita et al. 2013 for a computational model of how learners might solve this categorization problem). Second, they must be able to represent the hierarchical phrase-structure of the language, in order to identify the c-command relations between NPs. This structural knowledge too will depend on experience with the language, as different languages express hierarchical structures with different word orders. Finally, even if a child has successfully identified the pronouns and the hierarchical structure of the language, deploying that knowledge in real-time understanding will depend on the efficiency of the parsing mechanisms used in the process of sentence understanding.

Sentence processing research with adults reveals that c-command relations are used to guide the search for potential pronominal antecedents (Kush, Lidz & Phillips 2015, Cunnings, Patterson & Felser 2014) and that Principle C functions as an immediate constraint on pronominal interpretation (Kazanina et al. 2007, Kush, Lidz & Phillips 2017). Upon encountering a pronoun without an antecedent, adults will ignore any NPs in the c-command domain of the pronoun as candidate antecedents (Kazanina et al. 2007). These results suggest that adults deploy their grammatical knowledge accurately and immediately and that sentence

[1] There are also some well-known apparent counterexamples to Principle C (Evans 1980):

a. A: Is that Sarah? B: It must be. She's wearing Sarah's coat.
b. Everyone hates Oscar. Even he hates Oscar.Such cases reflect the fact that not every situation in which a pronoun and a name corefer is due to a binding relation between them. Since Principle C governs binding relations and not "accidental coreference", these cases do not constitute actual counterexamples to the generalization (for elaboration and tests to distinguish binding from accidental coreference, see Sag 1976, Williams 1977, Reinhart, 1983).

processing research can therefore function as a probe for that knowledge.

Principle C has received attention in language acquisition research for a number of reasons. First, the constraint is stable cross-linguistically; every language displays its effects, though in some languages these may be masked by independent features of the language (Baker, 1991, 2001). Further, work with 3- to 5-year-olds on Principle C has shown preschool-aged children to have robust knowledge of the constraint (for a review, see Lust, Eisele & Mazuka 1992). These characteristics have been considered by some to be "hallmarks of innateness," evidence that Principle C is specified as part of Universal Grammar (Crain, 1991). Whereas the innateness question is important, it is not our goal in this paper to contribute to this debate.

Lukyanenko et al. (2014) explored young children's understanding of binding constraints in a preferential looking task (Golinkoff et al., 1987) probing Principle C effects. They presented 30-month-olds with two events: a reflexive, one-participant action, where a character acted upon herself (e.g. *girl A patting herself*) and a non-reflexive, two-participant action, where another character acted upon the same girl (e.g. *girl B patting girl A*). When presented with a sentence like "*she's patting Katie,*" children were shown to preferentially attend to the two-participant image. Further, this result was shown to interact with children's vocabulary size: children with larger vocabularies looked more to the two-participant event in Principle C contexts than those with smaller vocabularies. The dispreference for a reflexive interpretation in Principle C contexts mirrors effects shown in older children and adults. This research is some of the first evidence that children younger than 3 years old reliably prefer the same restricted set of interpretations in binding environments as adults do.

However, it also presents a puzzle: There is no reason to expect a direct link between a child's vocabulary and their knowledge of Principle C, making it important to identify what explains the observed relation between vocabulary and Principle C. Is the relation between vocabulary size and Principle C explained by speed of lexical processing generally? Is vocabulary acting as an index of the child's syntactic knowledge? Or is vocabulary an index of pronoun familiarity or some aspect of pronoun processing that is independent of the structural knowledge underlying Principle C?

The goal of the current paper is to explore what aspects of children's developing linguistic abilities support their success in interpreting Principle C sentences. Because it makes reference to hierarchical structure, Principle C provides a clear probe of syntactic knowledge. And because this constraint is interpretive in nature, it is appropriate for testing with looking time measures that allow for precise, time-locked analysis (Fernald, Zangl, Portillo, & Marchman, 2008; Golinkoff et al., 1987). Consequently, we develop an independent test of children's abilities to build hierarchical structures in real time, what we call structural integration, in order to see whether performance in such a task predicts performance in a Principle C task.

Children's success in Principle C contexts, however, could be explained by their relying on non-hierarchical information, such as the linear relations between the noun phrases or the number of arguments in the clause, in order to arrive at their interpretations. Sutton (2015) eliminated the possibility that it is merely the number of arguments that cues children's interpretations. Thirty-month-olds show a preference for nonreflexive actions with sentences like "*she's patting Katie's head,*" in which the direct object noun phrase contains the name *Katie*, which is subject to Principle C, but they do not show such a preference for sentences like "*she's patting her head,*" in which the direct object noun phrase *her head* contains the pronoun *her*, which is not subject to Principle C (see also Sutton, Fetters, & Lidz, 2012). Nonetheless, there remain many non-syntactic strategies based in word order, such as a simple assumption that pronouns cannot precede their antecedents, that could yield the same response patterns. By exploring correlates of children's performance with these sentences, the research presented here aims to better understand the lexical and syntactic processing mechanisms underlying children's success.

In order to answer these questions, we examine correlations in performance between tasks that differ in their dependence on syntactic representations. By exploring which phenomena predict performance in Principle C contexts, we may be able to infer which types of processing mechanisms are involved in deriving an interpretation in these contexts.

Anticipating the results, we find no evidence that individual variation in performance on Principle C or reflexive sentences is predicted by variation in syntactic integration performance or lexical processing speed. We do, however, replicate the finding from Lukyanenko et al. (2014) that a child's vocabulary is a good predictor of performance on the Principle C task.

3. The present study

The research presented here explores predictors of individual variability in the interpretation of sentences subject to Principle C as a means of identifying the representations underlying this interpretation. We create an independent measure of structure-driven interpretation, along with a measure of lexical access speed, and ask whether these measures along with vocabulary are predictive of arriving at the correct interpretation in Principle C contexts. The basic line of argument is that if a particular underlying skill, such as structural integration or lexical access, is crucial to two tasks, we might expect to see correlations between those tasks.

We explore several factors that may be relevant to implementing interpretations. First, one of Lukyanenko et al.'s chief findings was that the size of children's productive vocabulary predicted performance in Principle C and reflexive contexts. What role does vocabulary play? No account predicts that vocabulary itself should directly affect the acquisition of Principle C[2]; it is unclear how the size of a child's lexicon would bear any direct relation to constraints on anaphora. For this reason, it seems that vocabulary may be the surface index of a different underlying mechanism (or mechanisms), for which variability more straightforwardly predicts variability in performance with Principle C.

Lukyanenko et al. (2014) suggest two possibilities. One is that vocabulary could reflect some aspect of children's grammatical development, such that the absence of Principle C effects in the low vocabulary children stems from their lacking some aspect of grammatical competence which allows successful application of the constraint. For example, if children with smaller vocabularies do not yet command the structure of transitive clauses, then they will be unable to successfully build a structure for the sentences and compute c-command relations to determine if binding conditions hold. The possibility that the vocabulary effect stems from different levels of grammatical development is supported by research showing vocabulary to be indicative of children's grammatical development (Dale, Dionne, Eley, & Plomin, 2000; Devescovi et al., 2005; Marchman & Bates, 1994). Alternatively, vocabulary may be an index of children's speed of processing, such that the failure of low vocabulary children to show Principle C effects in Lukyanenko et al. (2014) is a result of their inability to complete the interpretation and mapping processes quickly enough for this task. This possibility is supported by research in the word recognition domain, showing that vocabulary is related to children's speed of processing

[2] One exception would be that low vocabulary children could be predicted to fail on such a task if they do not know the verbs that were used in Lukyanenko et al.'s sentences; however this seems unlikely, both because the verbs used were highly common actions (cover, dry, fan, paint, pat, spin, squeeze, wash), and each action was introduced separately prior to the test phase.

(Fernald, Perfors, & Marchman, 2006 and Hurtado, Marchman, & Fernald, 2008, Weisleder & Fernald, 2013, inter alia).[3] A third possibility is that vocabulary is a reflection not of any syntactic knowledge or process, but rather of children's ability to deploy appropriate interpretive mechanisms upon encountering a pronoun. That is, children with lower vocabularies may be less proficient in identifying pronouns and deploying appropriate antecedent identification mechanisms, thus hiding any possible role for structure-dependent interpretation.

We therefore measure children's success in the Principle C task, along with three other aspects of their linguistic abilities, vocabulary, Lexical Access Speed (LAS) and Phrase Structure Integration (PSI), in order to determine whether these are predictive of children arriving at the correct interpretation in Principle C and reflexive sentences.

In several analyses we address the following questions:

Q1. Do children successfully distinguish interpretations in Principle C and reflexive contexts? Does there appear to be a vocabulary effect on success in the Principle C task? (i.e., Do the current results replicate the findings of Lukyanenko et al. (2014)?)

Q2. Do children successfully interpret hierarchical structure in our new task testing PSI?

Q3. To what extent do our measures of vocabulary, Lexical Access Speed and Phrase Structure Integration capture distinct aspects of children's developing linguistic abilities?

Q4. Which of these factors, if any, predict performance in Principle C and reflexive sentences? What inferences can we therefore make about the mechanism driving Principle C effects?

Thus, the goal of this study is to compare variability in various independent measures of linguistic ability against performance in Principle C contexts. If children's interpretations are structure dependent, then their processing of syntactic information should play a role in arriving at that interpretation, as it does in adults (Kazanina et al. 2007, Kush, Lidz & Phillips 2015, 2017, Cunnings, Patterson & Felser 2015). In contrast, if the bottleneck for children who do not succeed is in the ability to quickly access words from the lexicon and identify their meanings, we should instead see the LAS or vocabulary measures providing the strongest predictors of success. This research thus serves both the empirical goal of identifying the cause of early Principle C effects as well as the methodological goal of using measures of processing mechanisms to implicate specific interpretive mechanisms.

4. Materials and methods

4.1. Participants

We tested 64 English-speaking children (32 girls, 32 boys) 28–32 months of age (range = 28;0–31;27; median = 30;5; mean = 30;5) recruited through the University of Maryland Infant and Child Studies Database. Each participant completed three tasks: a Principle C task, a Lexical Access Speed task and a Phrase Structure Integration task, described below. Tasks were completed in one session that lasted around 30 min (including play breaks between tasks when needed). Five additional children were tested but were excluded from the final sample for the following reasons: failure to complete all three tasks ($n = 2$); equipment failure/experimenter error (n = 3).

MacArthur-Bates Communicative Development Inventory (MCDI) Words and Sentences long forms (Fenson et al., 1994) were collected for each child, revealing a range of parent-reported productive vocabulary from 99 to 689 words (median = 562; mean = 515).

4.2. Principle C task

The task designed to test interpretations in Principle C contexts was identical to the task in Lukyanenko et al. (2014), with one exception: Children were tested in a between-participants design, rather than receiving both Principle C and reflexive sentences in the same session.

In the stimulus videos, children were first presented with video clips introducing each of two characters (Katie and Anna), who would be performing the actions during test. Over the course of the task, children also received six identification trials, which presented these two characters on opposite sides of the screen, along with a sentence asking them to find one or the other (e.g. *Where's Katie? Do you see Katie?*). These trials ensured that the children were adequately mapping the names they heard in the introductory clips to the accompanying faces and could distinguish the two characters from one another.[4] Additionally, these trials also served to prepare children for the test trials, where they would be required to preferentially attend to one of two images on the screen. The target character and the side of the screen on which the target character appeared were counterbalanced across character identification trials and order of trials was counterbalanced across participants.

Actions represented in test trials were all continuous two-person scenes. The reflexive, one-participant actions consisted of scenes with one character performing an action on herself (e.g. Fig. 1 leftmost image, Katie patting her own head), with the other character present but not interacting. Two-participant, non-reflexive actions consisted of similar scenes, with one character performing the same action on the other character (e.g. Fig. 1 center image, Anna patting Katie's head). Appendix A presents a complete list of the stimuli used. Each test trial consisted of two phases: Familiarization and Test, presented in direct sequence. During the Familiarization phase, children saw both of the events that would be presented in the test phase, presented sequentially. Audio in each Familiarization phase clip, shown in (7a-b), described the action, but did not mention the agent or patient of the event by name.

(7) a. It looks like somebody is getting patted!

b. Somebody is getting patted again!

Order in which the Familiarization phase events appeared was counterbalanced across trials and participants. During the Test phase, children saw both of the events presented simultaneously on either side of the screen (e.g. Fig. 1, rightmost image). Audio consisted of three repetitions of the test sentence in slightly different frames, as shown in (8).

(8) a. She's patting Katie. Do you see the one where she's patting Katie? Find the one where she's patting Katie!

b. She's patting herself. Do you see the one where she's patting herself? Find the one where she's patting herself!

Children were presented with a total of 8 test trials in a between-participants design: half of the participants heard NAME condition sentences as in (8a) for all trials, and half heard all REFLEXIVE condition sentences as in (8b).

Across trials, sentences were aligned at the offset of the disambiguating object NP to simplify analysis. We measured the proportion of time children spent looking to the two-participant action in a 2000-ms window anchored 200 ms after the offset of the first iteration of the test sentence. If children are interpreting these sentences in adult-like ways,

[3] However, these findings are not entirely uncontroversial; several studies have found little or no evidence for a significant relation between vocabulary and processing speed (Swingley et al., 1999; Swingley & Aslin, 2000; Hurtado, Marchman, & Fernald, 2007). Even within studies, results can vary by age (Fernald et al., 2006; Hurtado et al., 2008).

[4] Strictly speaking, knowing the characters' names was not crucial to forming an interpretation. Thus these identification trials primarily served to facilitate processing of subsequent sentences.

Fig. 1. Schematic of Familiarization and Test phase sequence in Principle C task.

performance should differ in the NAME and REFLEXIVE conditions: proportion looking to the two-participant action should be high in the NAME condition, and low in the REFLEXIVE condition. Furthermore, if we replicate the patterns in Lukyanenko et al. (2014), this difference should be clear among children with larger vocabularies and slight or absent among children with smaller vocabularies.

The two remaining tasks were each designed to measure an additional factor that has been shown to relate to vocabulary size, and which might help to explain the vocabulary effect.

4.3. Lexical access speed task

The Lexical Access Speed (LAS) task, which generated measures of each child's lexical processing speed, was a word-object mapping task modeled after that of Swingley, Pinto, and Fernald (1999). Children were presented with two images of common objects, and a sentence which then directed children to find one of the two objects. Objects presented were chosen from the most common nouns in young children's vocabularies (all words, listed in Appendix B, are reported to be said by at least 90% of 30-month-old children). Fig. 2 presents a sample array. After observing the images in silence for approximately 1 s, children heard two instances of the test sentence, naming one of the two items, as in (9).

(9) Where's the train? Do you see the train?

Each of the 8 trials lasted a total of 5 s. Position of the target object was counterbalanced across trials; target object and order of presentation (2 possible lists) were counterbalanced across participants.

The speed measure standardly used in word-learning literature corresponds to the mean latency after the onset of the disambiguating word to re-orient to the target image on distractor-initial trials (Swingley et al., 1999; Weisleder & Fernald, 2013). Because a shift in attention is only appropriate on distractor-initial trials, it is from this subset of the data that the measure is derived. The point of disambiguation for each trial is defined as the point at which children have enough information to form an accurate interpretation of the sentence. In the LAS task, this point is the onset of the target noun.

Reaction time (RT) was therefore calculated by determining the latency to attend to the target image on distractor-initial trials (i.e., trials in which the child happened to be looking at the distractor image at the onset of the target word) in a 1500-ms window, anchored 200 ms after target noun onset. These RT values were then averaged across trials to derive a LAS measure for each participant.

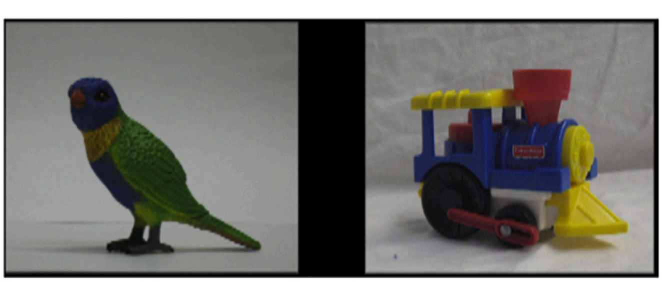

Fig. 2. Lexical access speed task sample array.

4.4. Phrase structure integration task

While the processing speed measure derived from the LAS task is widely utilized in word recognition research, a comparable standard measure of children's ability to build and integrate information over hierarchical structure does not yet exist; we have therefore designed a Phrase Structure Integration (PSI) task in order to generate such a measure. We developed a task in which the combination of linguistic stimulus and visual context requires children to compute the hierarchical structure of a phrase to identify the intended meaning, rather than being able to rely on lexical information alone. To accomplish this, we utilize constructions with both a superlative and a color adjective as in (10), with a corresponding visual array as in Fig. 3.

(10) Where's the biggest red train?

Given the visual array presented in Fig. 3, there is no one item that could be identified as the target by simply identifying the relevant features in absence of a structured representation. Consider how interpretation would occur if no internal structure was applied to the phrase, as with the flat structure demonstrated in Fig. 4.

If children do not attribute hierarchical structure to the phrase, then each element would be interpreted conjunctively, and the target item

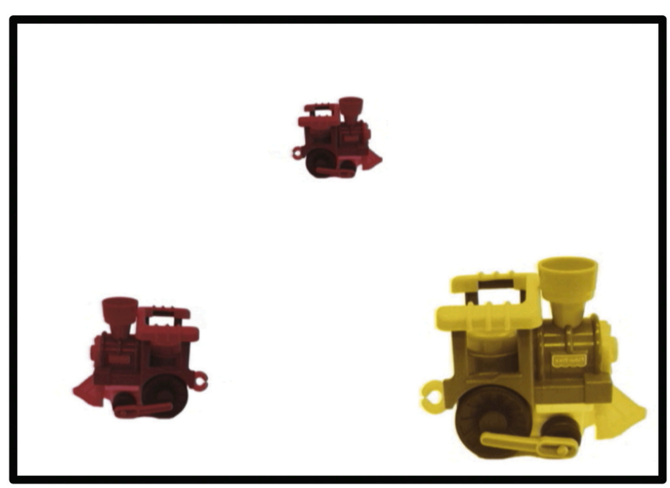

Fig. 3. Sample visual array for Phrase Structure Integration task. Largest item is yellow; medium and small items are red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Flat structure for "the biggest red train". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

would be identified as one which satisfies the combination *biggest* + *red* + *train*. None of the three items pictured satisfies all of the features *biggest* + *red* + *train*, because the item that is globally biggest does not satisfy the feature *red*, and the items that are red cannot be interpreted as satisfying the feature *biggest*.[5] Thus, in order to arrive at an adult-like interpretation of (10), children would need to represent the NP *biggest red train* hierarchically, as with the structure depicted in Fig. 5.

Given this hierarchical structure, the phrase [red train] can be interpreted as a unit to which the superlative *biggest* applies. With this interpretation, the biggest item in the set satisfied by the features *red* + *train* is the target. Thus with this superlative construction, we are able to identify a simple case in which accurate interpretation can be attributable to the use of hierarchical structure. In this way, children's success on this task can be considered a measure of processing and integrating syntactic information.

We maintained the word-object mapping task design in order to keep task demands as comparable as possible to the LAS task. Children were presented with three images of all the same kind (e.g. three trains); objects were drawn from the most commonly known nouns (e.g., trains, hats, books; see Appendix C). The three objects in each set varied in size

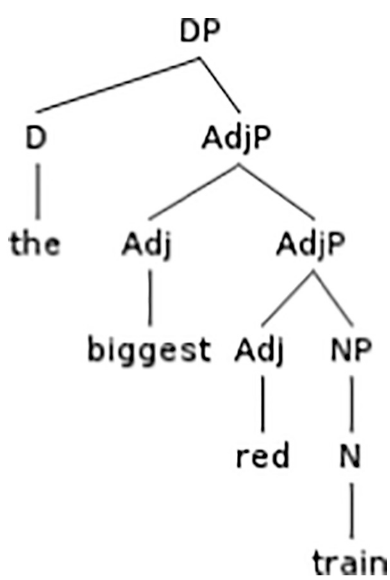

Fig. 5. Hierarchical structure for "the biggest red train". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and color, with the two smaller items sharing a color. Fig. 3 above presents a sample array. Each of 24 arrays contained objects of a different type. A constant size ratio of 3: 4.5: 7.5 was maintained between the smallest, medium, and largest item. This ratio was chosen so that the smallest item remained easily identifiable, so that the largest item was contained within its quadrant of the screen (to facilitate accurate coding), and so that the medium item was differentiable from the smallest item, yet sufficiently smaller than the largest item that it could not independently be considered big. After observing each array in silence for approximately 1 s, children heard an introductory sentence that identified the type of objects in the array (e.g. *Oh look! Now there are some trains!*).[6] Children were then presented with the test sentence. In 12 of 24 trials, the sentence contained the superlative *biggest* but no color adjective, as in (11a), indicating the globally largest item. In 12 trials, the sentence contained a color adjective, as in (11b), indicating the larger item in the subset of the two similarly colored items.

(11) a. Where's the biggest train? SUPERLATIVE

b. Where's the biggest red train? SUPERLATIVE + COLOR

SUPERLATIVE trials were 8.37 s long; SUPERLATIVE + COLOR trials were 8.77 s long. Order of item presentation was counterbalanced across participants. Position of the target item, color of the target and distractor items, and sentence type (SUPERLATIVE or SUPERLATIVE + COLOR) were counterbalanced across trials. Trials were presented in a pseudorandom order (or its reverse), chosen such that each of these factors was interspersed relatively evenly across the task.[7]

The superlative was a size adjective because this allowed the strict controlling of the ratio between the three objects in the array, and across trials, which is less feasible with other adjectives (e.g. specifying levels of fuzziness for *the fuzziest blue cat*). The particular size adjective *biggest* was used because *big* has been found to be the most frequently used base form for both comparatives and superlatives in young children's production (Layton & Stick, 1979). This coincides with data from the MCDI Lexical Norms Database showing *big* to be one of the earliest prominent adjectives in young children's expressive and receptive lexicons (Dale & Fenson, 1996). We avoided using *smallest* due to research showing that children acquire the positive dimensional adjectives before the corresponding negative adjective (Donaldson & Balfour, 1968; Ehri, 1977). The color words *red*, *blue*, *yellow*, and *green* were used due to their being the most common color words in children's productive vocabularies by this age (with the exception of *orange*, which was not used because of the potential confounding of the color name with the fruit).

Because size differences also correspond to differences in visual salience, we measured success in the task based on looking to each item in a 2000 ms pre-disambiguation window during the initial introductory sentence and a 2000 ms post-disambiguation window anchored 200 ms after the offset of *biggest* (i.e., 200 ms after the disambiguation point, which was the onset of either the noun or the color adjective), and Phrase Structure Integration as the difference between looking to the medium item during the post-disambiguation window in SUPERLATIVE + COLOR trials, where it was the target, and looking to the medium item in SUPERLATIVE trials, where it was not. Children who most readily process and integrate syntactic information should have larger PSI difference

5 Importantly, a norming study with adults indicated that the middle-sized train was not judged to be big, so even if children did not understand or chose to ignore the superlative suffix, there is no object that is big, red and a train, though there is something that is big for a red train (see Syrett & Lidz, 2010 for evidence of 30-month-olds understanding of relative adjectives).

6 Fernald, Thorpe, and Marchman (2010) showed high rates of 'false alarm' shifts to the distractor image when 30-month-old children were presented with two objects of the same kind and an adjective + noun phrase picking out one of the items (e.g. *where's the blue car* in the context of a blue and a red car). Therefore we included this introductory sentence to help ameliorate processing of the target noun.

7 The target item never occurred in the same position on consecutive trials. The same sentence type occurred in no more than two consecutive trials. The same color was presented in no more than three consecutive trials and presented in the same position in no more than two consecutive trials.

scores, looking more to the medium item when it is the target and less when it is not. For children who struggle more with integration and attempt to interpret the properties conjunctively, the large item (*biggest*) and the small item (color match) should be strong competitors, resulting in relatively few looks to the medium item in both SUPERLATIVE + COLOR and SUPERLATIVE trials, and correspondingly smaller difference scores.

4.5. Apparatus and procedure

Participants were tested individually, sitting either in a high chair or on their parent's lap in front of a 51-in. plasma television. A camera mounted above the television recorded participants' eye movements during the videos. Data was coded frame-by-frame using SuperCoder software (Hollich, 2005), indicating, in the Principle C and Lexical Access Speed tasks, whether children were attending to the left or right side of the screen, or not at all. Coders were trained researchers who were unaware of participants' assigned condition and could not hear the auditory stimuli. Five percent of the data was coded by all three coders, to ensure accurate coding and reliability across coders. Inter-coder reliability was high: across three coders, agreement was above 96% in all cases, with Cohen's kappa scores of 0.94 and above. For the PSI task, slight differences in coding were required. The three-way looking design necessitated coding whether children were looking to the right, left, and center portions of the screen, or not at all. Inter-coder reliability remained high; across three coders, agreement was above 97% in all cases, with Cohen's kappa scores of 0.95 and above.

5. Results and discussion

We first present results from each of the three tasks individually, and then examine to what extent participants' vocabulary, LAS or PSI measures successfully predict their behavior in the Principle C task.

5.1. Principle C task

Analysis of overall performance on the Principle C task assess whether children in the current study successfully distinguished interpretations in Principle C and reflexive contexts and whether children's performance in Principle C and reflexive contexts varies with vocabulary, as in Lukyanenko et al. (2014).

Fig. 6[8] shows children's patterns of looking to the two-participant video for both conditions. Before the point of disambiguation (the first dotted line), children in both conditions look about equally to the two-participant event. After hearing the disambiguating object NP, performance diverges: Participants in the NAME condition look more at the two-participant event than participants in the REFLEXIVE condition do. This suggests that children are interpreting both the reflexive and the Principle C sentences in relatively adult-like ways, and strongly resembles the results from Lukyanenko et al., despite the current study's between-participants design.

To assess this more precisely, we analyzed a 2000 ms window beginning 200 ms after the offset of the disambiguating object NP in the first repetition of the test sentence (as indicated in Fig. 6 by the solid box). Trials were excluded if participants spent more than 75% of the target window looking away. This resulted in the exclusion of 14 of 512 trials (2.7%), leaving 498 for analysis. We calculated the average proportion of time each participant spent looking to the two-participant video in the target window. Participants in the NAME condition looked reliably more at the two-participant action than participants in the REFLEXIVE condition in this window ($t(62) = 3.94$, $p = .0002$, two-tailed). This confirms that by 30 months, children look preferentially to an image depicting a disjoint rather than a coreferential interpretation in Principle C contexts, consistent with a constraint on pronoun interpretation.

To further explore the vocabulary effect observed by Lukyanenko et al., we analyzed performance by comparing behavior based on MCDI vocabulary. Lukyanenko et al. observed that below the median MCDI vocabulary (504.5 words), vocabulary did not predict performance, while for the high-vocabulary children in the NAME condition, knowing more words predicted increased looking to the two-participant event. Additionally, a parallel pattern emerged for the REFLEXIVE condition: larger MCDI vocabulary predicted increased looking to the one-participant event.

Fig. 7 shows the performance in each condition of the Principle C task by MCDI vocabulary in both the Lukyanenko et al. (2014) data and the current study. In both conditions and in both studies, performance across the lower half of the vocabulary range hovers around 50%. It's only in the upper half of the vocabulary range that we observe participants whose scores are strongly influenced by the sentences they heard (i.e., particularly high in the NAME condition and low in the REFLEXIVE condition). The close match between studies suggests that Lukyanenko et al.'s vocabulary effect was neither an artefact of their within-participants design nor of their much longer test window (9 s).

5.2. Lexical access speed task

Fig. 8 shows children's overall performance on the Lexical Access Speed task. Children were overwhelmingly successful in attending to the target object upon hearing the target word, as indicated by the sharp increase in the proportion of looking to the target after the onset of the target word (the zero-point on the x-axis). By 30 months, this type of task is exceedingly easy for children in general, but we expect differences in the speed with which children orient to the target image to be indicative of individual variation in speed of lexical processing.

The LAS measure was each child's mean latency to shift to the target image on distractor-initial trials, measured in a 1500 ms window (marked by the box in Fig. 8) anchored 200 ms after the onset of the target word.[9] Of 512 total trials, 7 were excluded for >75% away looks during the analysis window (1.4%). Of the remining 505 trials, 262 (51.9%) were distractor-initial. Of those, 198 contained direct distractor to target shifts within the 1500 ms analysis window. Participants contributed a median of 3 RTs (range 1–6), and the individual mean LAS measures ranged from 267 to 1200 ms (Median = 521 ms, Mean = 548 ms). This suggests that at 30 months, children's speed of processing at the lexical level is still quite varied.

5.3. Phrase structure integration task

Before using the Phrase Structure Integration measure as a potential predictor of performance in the Principle C task, we must first establish that children successfully interpreted hierarchical structure in this task. Fig. 9 shows the time course of children's looking in the Phrase Structure Integration task by condition. Because the visual array in this task consists of a target item and two distractors and looks therefore do not trade off in direct proportion to each other as looks to two items do, these graphs show the proportion of looks to each of the three items. Recall that in the SUPERLATIVE condition, where children hear sentences like *where's the biggest train*, the target item will be the large item, while in the SUPERLATIVE + COLOR condition, where children hear sentences like *where's the biggest red train*, the target will be the medium item (corresponding to the larger of the two similarly colored items). One obvious pattern in both conditions is an overall bias to look at the largest item before the disambiguation point. The two conditions differ dramatically

[8] All graphs were generated using the ggplot2 library (version 3.3.3, Wickham, 2016) in R (version 4.0.3, R Development Core Team, 2020).

[9] Shifts occurring less than 200 ms following the disambiguation point are not included in these calculations, to account for the time required for saccade programming.

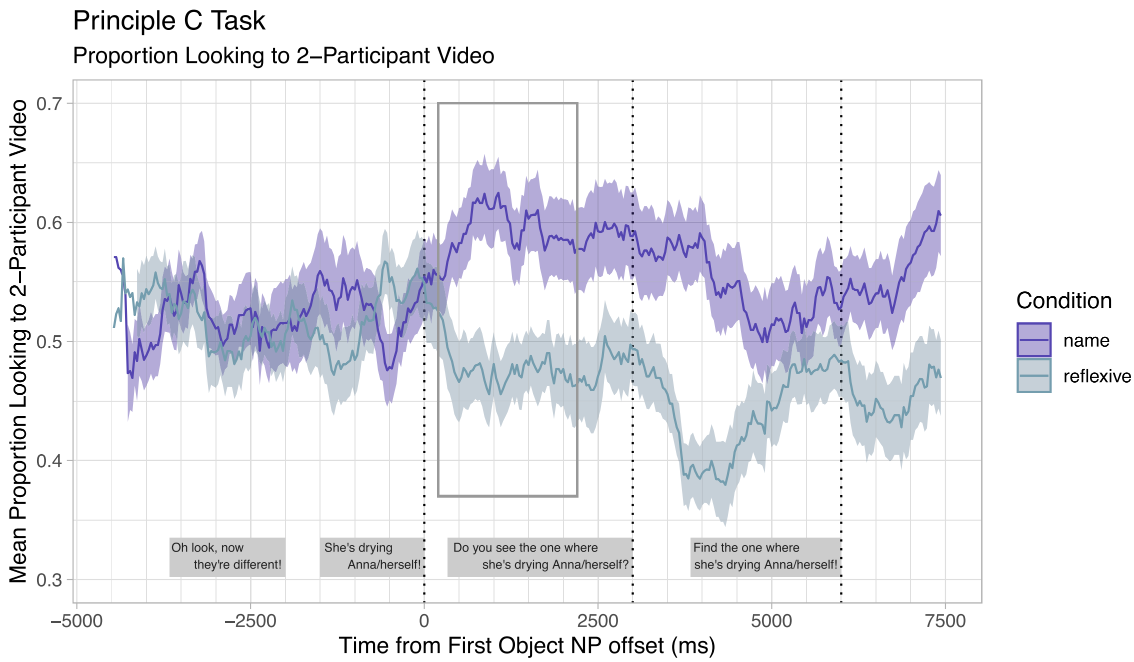

Fig. 6. Mean proportion looking to the two-participant action in the Principle C Task. Shaded areas show standard error of the mean. The outlined box represents the critical window for analysis (2000 ms beginning 200 ms following the offset of the disambiguating object NP in first iteration of the test sentence). The dotted lines represent the offset of the disambiguating object NP in each iteration of the test sentence. The NAME condition is shown in the darker blue and the REFLEXIVE condition is shown in the lighter blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

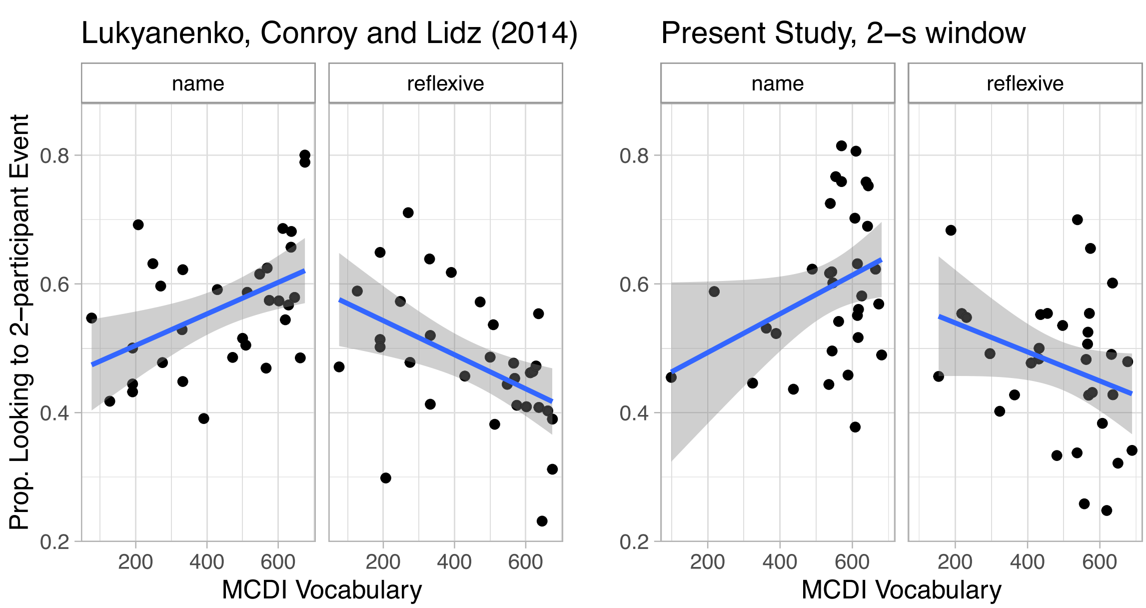

Fig. 7. Looking behavior in Principle C task by MCDI Vocabulary Score.

post-disambiguation, however. In the SUPERLATIVE condition, children continue looking most at the largest item, and in the SUPERLATIVE + COLOR condition, looks to the largest item drop off and are overtaken by looks to the medium item, the target.

Despite the dramatic difference between conditions in the post-disambiguation window, visual inspection of the patterns of looking over time is not sufficient to confirm that children succeeded at Phrase Structure Integration in the SUPERLATIVE + COLOR condition. Looks to both the medium and the small item increase after children hear the color adjective, as one might expect if children are interpreting the two adjectives conjunctively or in sequence rather than integrating them.

To confirm that children successfully integrated the two adjectives, we compared children's looking to each item in pre- and post-disambiguation windows for each type of trial. Windows were excluded if children looked away for more than 75% of the frames ($N = 424$ of 3072 windows, 13.8%). This left 2648 windows for analysis. A summary of the resulting proportions is shown in Fig. 10. Average proportion looking to the largest item increased in both conditions, though more dramatically in the SUPERLATIVE condition than in the SUPERLATIVE + COLOR condition, and looking to the smallest item decreased in both conditions. In contrast, looks to the medium item decreased in the SUPERLATIVE condition, and increased in the SUPERLATIVE + COLOR condition.

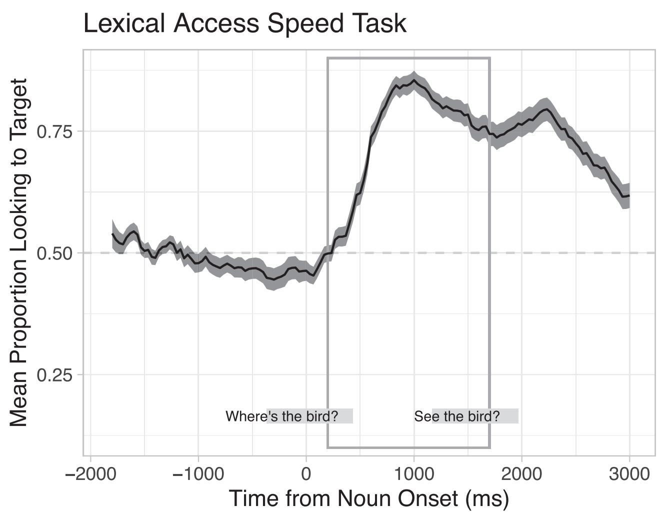

Fig. 8. Mean proportion looking to the target in Lexical Access Speed Task. The shaded area represents the standard error of the mean. The outlined box represents the critical window for RT calculation.

structure of the Noun Phrase.

To test these patterns, proportion looking to the relevant item by window and condition was analyzed in a series of three linear mixed effects models.[10] Condition and window were within-participant categorical predictor variables and were effects coded (condition: SUPERLATIVE −0.5, SUPERLATIVE + COLOR 0.5; window: pre −0.5, post 0.5). The models included random intercepts by item and participant. Both the model for looking at the large item and the model for looking at the small item revealed reliable main effects of window (large: $N_{obs} = 2648$, $b = 0.04$, $se = 0.01$, $t = 3.41$, $p = .0007$; small: $N_{obs} = 2648$, $b = -0.06$, $se = 0.01$, $t = -5.50$, $p < .0001$),[11] though these effects go in opposite directions. Looking increased to the large item and decreased to the small item. In contrast, the model for looking at the medium item showed a reliable main effect of window ($N_{obs} = 2648$, $b = 0.02$, $se = 0.01$, $t = 2.07$, $p = .04$), and interaction between window and condition ($b = 0.05$, $se = 0.02$, $t = 2.36$, $p = .02$). These results support the conclusion that, as a group, 30-month-olds successfully integrated the superlative and color adjective in the SUPERLATIVE + COLOR condition, increasing their looking to the medium item in the SUPERLATIVE + COLOR condition and decreasing it in the SUPERLATIVE condition.

To derive a PSI measure from this task, we compared looking to the medium item in the post-disambiguation window in the two conditions.

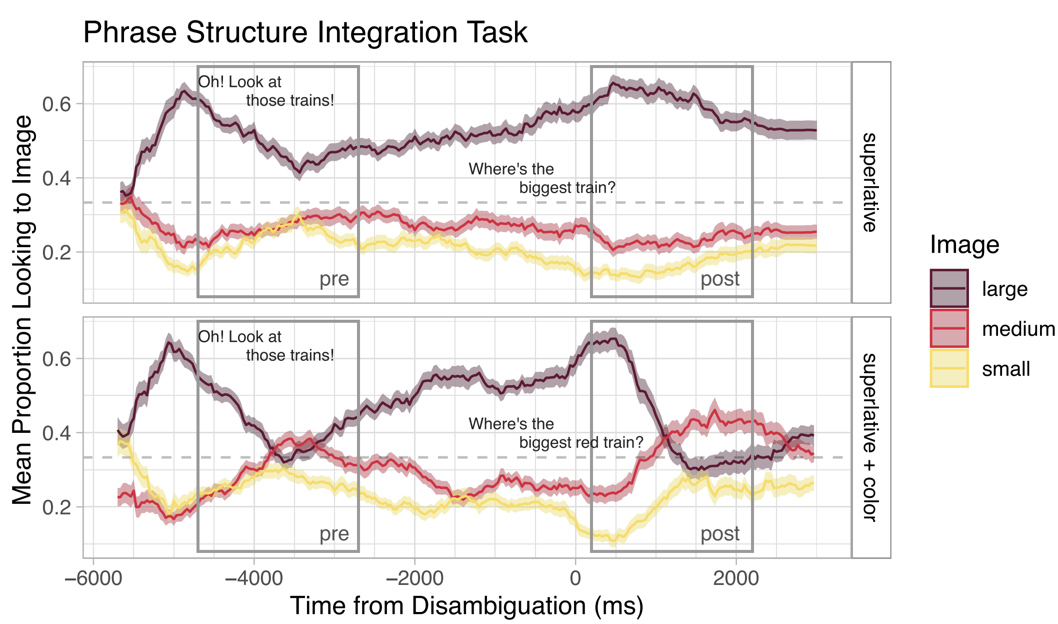

Fig. 9. Mean proportion looking to each item in Phrase Structure Integration Task by condition. The shaded areas represent the standard error of the mean. Outlined boxes represents critical windows for analysis: analyses of success in the PSI task use both the pre- and post-disambiguation windows, the PSI measure uses only the post-disambiguation window.

These patterns suggest that children successfully integrated the superlative and color adjective in the SUPERLATIVE + COLOR condition, rather than interpreting them in sequence. In particular, the fact that looks to the medium item but not the small item increase from the pre- to the post-disambiguation window in the SUPERLATIVE + COLOR condition suggests that children were not simply increasing their looking to color-matching items. Similarly, the fact that looks to the medium item increase in the SUPERLATIVE + COLOR condition and decrease in the SUPERLATIVE condition indicates that increasing looks to the medium item are not due to, e.g., boredom with the initially attention-grabbing largest item. In this task, accurate interpretation and selection of the target object requires children to treat the test sentences as hierarchically structured phrases. Thus these patterns suggest that children are able to select the target item in both conditions, and therefore that at 30 months, children's interpretations are influenced by the internal hierarchical Children who are most successful at integrating the superlative and color adjectives should show the largest differences between looks to the medium item when it is the target and when it is not. Participants' PSI

[10] All mixed effects models were fit using the lme4 package (version 1.1–26, Bates, Mächler, Bolker & Walker, 2015) in R (version 4.0.3), and *p*-values were calculated by lmerTest, using Satterthwaite denominator degrees of freedom (3.1–3, Kuznetsova, Brockhoff, & Christensen, 2017).

[11] Equivalent models of empirical logit transformed proportion looking to each item were analyzed in the same way. The direction and significance level of all effects were the same in the analyses of transformed and untransformed proportions, with one exception: in the transformed model of looks to the large item, the interaction of window and condition is reliable ($b = -0.4$, $se = 0.2$, $t = -2.52$, $p = .012$). Untransformed proportions are reported here for ease of interpretation.

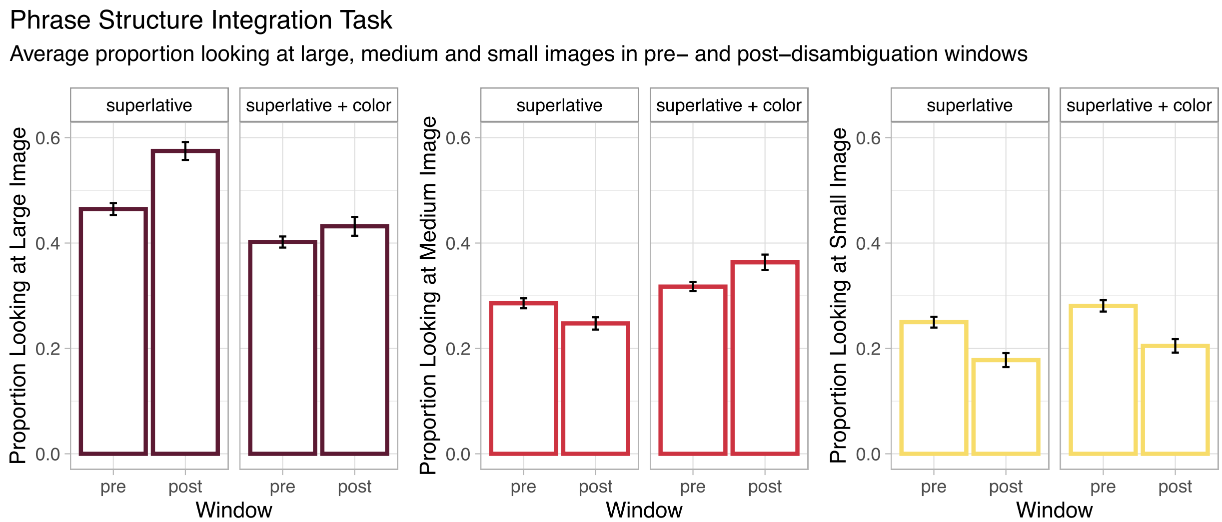

Fig. 10. Average proportion looking time to each item in the PSI task by window and condition. Error bars show standard error of the mean.

scores ranged from −0.130 to 0.502 (Mean = 0.116, Median = 0.116).

5.4. Comparison of covariate measures

Before asking whether vocabulary, Lexical Access Speed, and Phrase Structure Integration scores predict 30-month-olds' performance in the Principle C task, we explored the relationships between them. This allowed us to confirm that they are relatively independent of one another. None of these measures were significantly correlated with any others (LAS-vocabulary $r(62) = -0.12$, $t = -0.98$, $p = .33$; PSI-vocabulary $r(62) = 0.18$, $t = 1.45$, $p = .15$; PSI-LAS $r(62) = -0.11$, $t = -0.84$, $p = .4$), suggesting that they are capturing different aspects of children's linguistic abilities. This data is plotted in Fig. S-1 in the supplemental materials.

Further, two points of evidence here allow us to draw some interim conclusions related to our question of how vocabulary plays a role in children's interpretations in Principle C contexts. First, the lack of reliable correlation between MCDI vocabulary and LAS suggests that processing speed is not a reliable predictor of vocabulary in 30-month-olds, at least for this group of participants.[12] Second, the lack of correlation between MCDI vocabulary and the PSI measure suggests that syntactic integration is also not a reliable predictor of vocabulary at this age. Together, these points weaken the argument that the vocabulary effect on children's Principle C knowledge is a reflection of variation in processing speed or of syntactic development. It is worth noting, however, that despite the lack of significant relationships, all correlations go in the predicted directions: PSI and vocabulary are positively correlated, and LAS is negatively correlated with the other measures. Due to the lack of a relation shown between any of these measures, in the following analyses we treat each as an independent covariate measure.

[12] Of course, it is possible that 30 months is simply beyond the age range when vocabulary can be reliably linked to lexical processing speed. This could be due to either ceiling effects on the vocabulary measure, or to the fact that there is less variability in lexical processing than there is at younger ages. Additionally, one could raise the methodological issue that tasks measuring processing speed in the literature generally use more trials; while our task included 8 trials, most tasks in previous research have used upwards of 20 trials. Ongoing research aims to further explore the connection between vocabulary and various measures of processing speed with increased number of trials to facilitate having enough distractor-initial trials to form reliable measures.

5.5. Predicting success with principle C

The results in this section address the final question posed in Section 3: which of the identified factors (vocabulary, lexical processing speed, and syntactic processing speed), if any, predict performance in Principle C contexts? What inferences can we therefore make about the mechanism driving Principle C effects?

So far we have demonstrated that children exhibit behavior consistent with knowledge of Principle C and reflexives at 30 months. Our analysis of individual differences in processing seeks to determine what type of processing is implicated in this response, as a means to determine the underlying knowledge behind children's performance. If successful interpretation is predicted by syntactic integration, then we can infer that the underlying knowledge driving behavior is structural in nature; this dependency is predicted if children are using accurate knowledge of Principle C. Alternatively, if successful interpretation is predicted by speed of processing at the lexical level, then we can infer that the underlying knowledge driving behavior is non-structural in nature. Finally, if vocabulary is the only predictor of performance on Principle C, this would suggest that vocabulary is indexing an independent feature of pronoun processing that is unrelated to the structure of the sentence.

To understand the best predictors of performance in the Principle C task, we fit a mixed-effects linear model of children's proportion looking to the two-participant action. The model included the between-participants categorical predictor condition (effects coded: NAME 0.5, REFLEXIVE −0.5), and continuous predictors for vocabulary, LAS, and PSI. The model also included the interactions of the continuous predictors with condition and random intercepts for participant and item. This model revealed a reliable interaction between condition and vocabulary ($b = 0.0005$, $se = 0.0002$, $t = 2.24$, $p = .025$).[13] No other predictors contributed reliably (condition $b = -0.27$, $se = 0.19$, $t = -1.43$, $p = .15$; all other $|t|$s. < 1, $ps > 0.4$). Follow-up comparisons using treatment coding revealed a marginal simple main effect of vocabulary in the NAME condition ($b = 0.0003$, $se = 0.0002$, $t = 1.71$, $p = .09$).[14] No other simple main effects were reliable (REFLEXIVE condition vocabulary $b = -0.0002$,

[13] In an identical model fit over empirical logit transformed proportions the interaction of condition and vocabulary was marginal ($b = 0.004$, $se = 0.002$, $t = 1.96$, $p = .05$). All other effects were non-significant in both models.

[14] The only difference between this model and the treatment coded models of empirical logit transformed proportions is that the marginal simple main effect of vocabulary in the NAME condition was non-significant ($b = 0.002$, $se = 0.002$, $t = 1.33$, $p = .19$). The simple main effect of vocabulary in the REFLEXIVE condition remained non-significant ($b = -0.002$, $se = 0.001$, $t = -1.46$, $p = .15$).

$se = 0.0002$, $t = -1.45$, $p = .15$; all other $|t|s < 1$, $ps > 0.4$). These results indicate that vocabulary is a more reliable predictor of success in the Principle C task than either LAS or PSI measures, and moreover, that it is a reliable predictor of success beyond any variance those measures account for. These patterns are consistent with the patterns shown in Fig. 11.

6. General discussion

The research presented here shows several key findings. Using a 2000 ms window of analysis, we find that 30-month-olds as a group show success in interpreting sentences in both Principle C and reflexive contexts. Replicating Lukyanenko et al. (2014), we find that vocabulary size is a reliable predictor of performance in both Principle C and reflexive contexts.

In the analysis of covariate measures, we find no correlations between our measures of vocabulary, lexical access speed and syntactic integration, suggesting that these are independent contributors to language understanding. Finally, we find no evidence that either lexical access speed or syntactic integration explains children's performance with Principle C. Thus, if we take these results at face value, we might conclude that the vocabulary is related to some aspect of interpretation that is tied neither to lexical access nor to syntactic integration, such as pronoun familiarity or antecedent identification.

Another important contribution of this work, though not the main thrust of our investigation, is the development of a novel method for measuring syntactic integration and the construction of structure-dependent interpretation inside the Noun Phrase. Here, we found that 30-month-olds successfully integrate a pair of adjectives in order to restrict the interpretation of the Noun Phrase containing them, highlighting a role for syntactic structure in guiding interpretation.

At the same time, this investigation is marked by a failure to find the precise mechanisms underlying children's success with Principle C. Whereas we successfully replicate the role of vocabulary as a predictor of children's interpretations, we were unable to identify the specific mechanisms that explain why vocabulary is related to interpretation. We fail to find an effect of lexical processing speed, suggesting that the vocabulary effect may index some other component of the ultimate interpretation. We similarly fail to find an effect of syntactic integration, suggesting that the vocabulary effect may not index hierarchical structure building. The most straightforward interpretation of the data, then is that our vocabulary effect indexes some aspect of pronoun interpretation that is not tied to either lexical access or syntactic structure. One possibility is that vocabulary size indexes overall knowledge of pronouns and a child's facility with engaging either the discourse representations associated with pronouns or the specific processing mechanisms that the pronominal interpretation requires. This hypothesis is potentially supported by the observation that performance in both reflexive and Principle C contexts is predicted by vocabulary size.

Nonetheless, we would caution against taking the results of this study as indicating that there is no role for structure-dependent interpretation in children's behavior with respect to Principle C, for both methodological and theoretical reasons. Methodologically, it is a truism that one should not draw strong conclusions from null results, but more importantly, finding correlations between reliable tasks can be very difficult, even given the medium sample sizes used here, because reliable effects are generally reliable because they exhibit relatively little variability (Hedge et al., 2017).

Theoretically, one strong possibility is that the kind of structure-integration mechanisms required to interpret complex Noun Phrases like *the biggest red train* are distinct from those required to deploy Principle C. Specifically, it is possible that an interpretive revision is required in successfully interpreting these Noun Phrases in our phrase structure integration task. A listener who, upon hearing "the biggest," commits to an interpretation that the NP will be "the biggest train" will have to revise that commitment upon hearing "red". Thus it is possible that variability in our syntactic integration task, at least in part, measures the revision abilities associated with conflict monitoring and cognitive control (Botvinick et al., 2001, Novick et al., 2005), thus hiding the potential contribution of structure building itself. We thus take it to be an important open question whether children's success with Principle C is a reflection of their syntactic knowledge. Further work will have to find ways of assessing the role of syntax in guiding interpretation at this age.

7. Conclusion

Drawing inferences about the nature of underlying syntactic knowledge can be challenging because observed performance is a reflection not only of this knowledge, but also of the deployment processes required to implement this knowledge in real time. Instead of treating performance effects as something to abstract away from, in this work we aimed to capitalize on this rich source of information by pairing tasks that require the same linguistic resources as a way to probe syntactic knowledge. We identified three possible factors that could contribute to interpretation in Principle C contexts: vocabulary, processing of lexical information, and processing of syntactic information. Our results show that performance on Principle C contexts is predicted by a child's vocabulary size, but not lexical access speed or our measure of syntactic integration. This increases the likelihood that early variability in success with Principle C may be constrained by a different correlate of vocabulary, such as pronoun familiarity or skill in antecedent identification. We did not succeed in identifying a clear role for

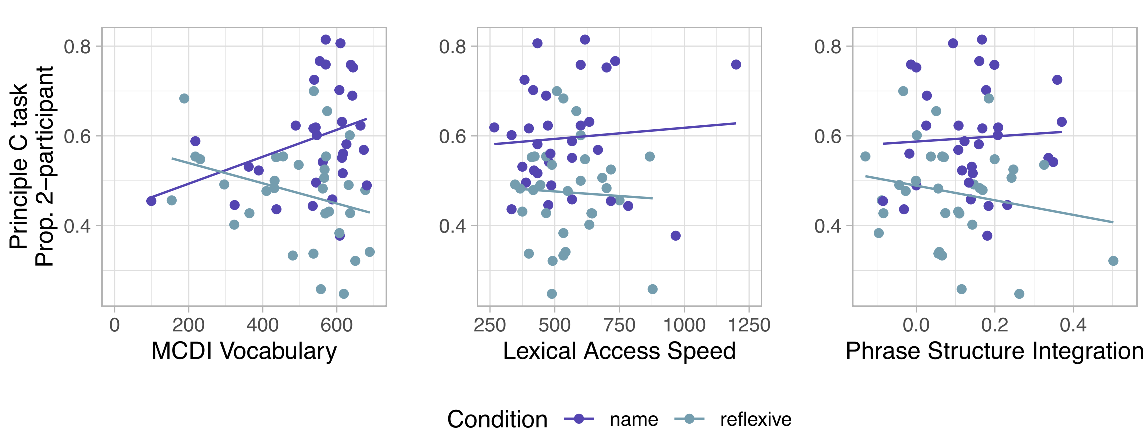

Fig. 11. Scatter plots showing the relationship between performance in the Principle C task and each of the covariate measures. Color represents condition.

syntactic processing in explain children's performance in Principle C. Nonetheless, we hope that this work opens up a new approach to studying young children's syntactic knowledge, capitalizing on the shared processing mechanisms across distinct tasks as a way of probing for shared structural representations.

Appendix A. Principle C task

Table A1
Schematic of Task.

Phase	Video		Audio
Filler	abstract		*classical music*
Character Intro	Katie		*Oh look! There's Katie. Katie's standing.*
Character Intro	Katie		*Oh wow! Look at what Katie's doing now.* *Katie's waving! Look at Katie waving.*
Character Intro	Anna		*Oh, there's Anna! Anna's standing too.*
Character Intro	Anna		*Look! Look at what Anna's doing now.* *Anna's waving! Look at Anna waving.*
Character Intro	Anna		*Wow, there's Anna again.*
Character Intro	Katie		*Oh look- there's Katie!*
Face Check 1	Anna	Katie	*Wow- there they are!* *Do you see Katie? Where's Katie?*
Character Intro	Katie		*Yay! There's Katie dancing.*
Face Check 2	Anna	Katie	*There they are again!* *Do you see Anna? Where's Anna?*
Character Intro	Anna		*Yay! There's Anna stretching.*
Familiarization 1	Anna dries Anna		*Oh wow- there's Katie and Anna!* *It looks like somebody is getting dried!*
	Katie dries Anna		*Hey look- there they are again!* *Somebody is getting dried again!*
Test 1	Anna dries Anna	Katie dries Anna	*Oh look! Now they're different. She's drying Anna. Do you see the one where she's drying Anna? Find the one where she's drying Anna.*
Face Check 3	Katie	Anna	*Oh look- they're jumping!* *Do you see Katie? Where's Katie?*
Filler	abstract		*classical music*
Face Check 4	Katie	Anna	*Oh look- they're jumping again!* *Do you see Anna? Where's Anna?*
Familiarization 2	Anna pats Katie		*Oh wow- there's Anna and Katie!* *It looks like somebody is getting patted!*
	Katie pats Katie		*Hey look- there they are again!* *Somebody is getting patted again!*
Test 2	Anna pats Katie	Katie pats Katie	*Oh look! Now they're different. She's patting Katie. Do you see the one where she's patting Katie? Find the one where she's patting Katie.*
Character Intro	Katie		*Hey- there's Katie marching!* *Do you see her marching?*
Familiarization 3	Katie paints Anna		*Oh wow- there's Katie and Anna!* *It looks like somebody is getting painted!*
	Anna paints Anna		*Hey look- there they are again!* *Somebody is getting painted again!*
Test 3	Katie paints Anna	Anna paints Anna	*Oh look! Now they're different. She's painting Anna. Do you see the one where she's painting Anna? Find the one where she's painting Anna.*
Character Intro	Anna		*Wow- there's Anna hopping!* *Do you see her hopping?*
Familiarization 4	Anna fans Anna		*Oh wow- there's Katie and Anna!* *It looks like somebody is getting fanned!*
	Katie fans Anna		*Hey look- there they are again!* *Somebody is getting fanned again!*
Test 4	Anna fans Anna	Katie fans Anna	*Oh look! Now they're different. She's fanning Anna. Do you see the one where she's fanning Anna? Find the one where she's fanning Anna.*
Face Check 5	Anna	Katie	*Look- there they are again!* *Do you see Anna? Where's Anna?*
Familiarization 5	Anna washes Katie		*Oh wow- there's Anna and Katie!* *It looks like somebody is getting washed!*
	Katie washes Katie		*Hey look- there they are again!* *Somebody is getting washed again!*
Test 5	Anna washes Katie	Katie washes Katie	*Oh look! Now they're different. She's washing Katie. Do you see the one where she's washing Katie? Find the one where she's washing Katie.*
Filler	abstract		*classical music*
Character Intro	Anna		*Hey- there's Anna stretching!* *Do you see her stretching?*
Familiarization 6	Anna spins Anna		*Oh wow- there's Katie and Anna!* *It looks like somebody is getting spun!*

(*continued on next page*)

Table A1 (*continued*)

Phase	Video		Audio
	Katie spins Anna		*Hey look- there they are again!* *Somebody is getting spun again!*
Test 6	Anna spins Anna	Katie spins Anna	*Oh look! Now they're different. She's spinning Anna. Do you see the one where she's spinning Anna? Find the one where she's spinning Anna.*
Character Intro	Katie		*Wow- there's Katie dancing!* *Do you see her dancing?*
Familiarization 7	Katie squeezes Katie		*Oh wow- there's Anna and Katie!* *It looks like somebody is getting squeezed!*
	Anna squeezes Katie		*Hey look- there they are again!* *Somebody is getting squeezed again!*
Test 7	Katie squeezes Katie	Anna squeezes Katie	*Oh look! Now they're different. She's squeezing Katie. Do you see the one where she's squeezing Katie? Find the one where she's squeezing Katie.*
Face Check 6	Anna	Katie	*Look- now they're waving!* *Do you see Katie? Where's Katie?*
Familiarization 8	Anna covers Katie		*Oh wow- there's Anna and Katie!* *It looks like somebody is getting covered!*
	Katie covers Katie		*Hey look- there they are again!* *Somebody is getting covered again!*
Test 8	Anna covers Katie	Katie covers Katie	*Oh look! Now they're different. She's covering Katie. Do you see the one where she's covering Katie? Find the one where she's covering Katie.*
Filler	abstract		*classical music*

Appendix B. Lexical access speed task

Table B1
Schematic of lexical access speed task.

Trial	Video		Audio
1	bird	train	*Where's the bird?* *See the bird?*
2	spoon	keys	*Where are the keys?* *See the keys?*
3	book	cup	*Where's the cup?* *See the cup?*
4	shoe	hat	*Where's the shoe?* *See the hat?*
5	ball	flower	*Where's the flower?* *See the flower?*
6	dog	cat	*Where's the dog?* *See the dog?*
7	horse	chair	*Where's the horse?* *See the chair?*
8	cookie	fish	*Where's the fish?* *See the fish?*

Appendix C. Phrase structure integration speed task

Table C1
Schematic of phrase structure integration speed task.

Trial	Video		Audio
Filler	abstract		classical music
1	large green		*Hey look- there are some shirts!*
	small red	medium red	*Where's the biggest shirt?*
2	medium yellow		*Oh- do you see the boats?*
	small yellow	large blue	*Where's the biggest boat?*
3	small yellow		*Oh hey- look at those cats!*
	medium yellow	large blue	*Where's the biggest yellow cat?*
4	medium red		*Hey look- there are some chairs!*
	small red	large green	*Where's the biggest red chair?*
5	small yellow		*Oh wow- look at those hands!*
	medium yellow	large green	*Where's the biggest hand?*
6	medium red		*Now there are some bikes!*
	large blue	small red	*Where's the biggest red bike?*
7	small green		*Oh hey- look at those books!*
	large blue	medium green	*Where's the biggest book?*
8	large red		*Wow- look at those hats!*
	medium yellow	small yellow	*Where's the biggest hat?*

(*continued on next page*)

Table C1 (*continued*)

Trial	Video		Audio
Filler	abstract		classical music
9	small green		*Now there are some shoes!*
	large red	medium green	*Where's the biggest green shoe?*
10	large yellow		*Oh wow- look at those houses!*
	medium green	small green	*Where's the biggest house?*
11	small blue		*Hey- look at those cars!*
	large green	medium blue	*Where's the biggest car?*
12	medium yellow		*Wow- look at those boxes!*
	large red	small yellow	*Where's the biggest yellow box?*
13	large yellow		*Now there are some cups!*
	medium green	small green	*Where's the biggest green cup?*
14	medium red		*Hey look- do you see those trucks?*
	small red	large blue	*Where's the biggest truck?*
15	large green		*Oh hey- look at the bears!*
	medium yellow	small yellow	*Where's the biggest yellow bear?*
16	small blue		*Hey look- do you see those dogs?*
	large yellow	medium blue	*Where's the biggest blue dog?*
Filler	abstract		classical music
17	medium blue		*Oh wow- look at those buses!*
	large red	small blue	*Where's the biggest bus?*
18	large green		*Oh hey- look at the plates!*
	small blue	medium blue	*Where's the biggest blue plate?*
19	small red		*Oh- look at those trains!*
	medium red	large yellow	*Where's the biggest red train?*
20	small blue		*Wow- now there are some balls!*
	medium blue	large yellow	*Where's the biggest ball?*
21	medium blue		*Hey look- do you see the horses?*
	small blue	large red	*Where's the biggest blue horse?*
22	medium green		*Oh- look at those dolls!*
	large red	small green	*Where's the biggest doll?*
23	large yellow		*Wow- now there are some blocks!*
	small red	medium red	*Where's the biggest block?*
24	large blue		*Oh wow- look at those bowls!*
	small green	medium green	*Where's the biggest green bowl?*
Filler	abstract		classical music

Appendix D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104676.

References

Baker, M. (1991). On some subject/object non-asymmetries in Mohawk. *Natural Language & Linguistic Theory, 9*, 537–576.

Baker, M. (2001). The natures of nonconfigurationality. In M. Baltin, & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 407–438). Oxford: Blackwell.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed Models Using lme4. *Journal of Statistical Software, 67*, 1–48.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001 Jul). Conflict monitoring and cognitive control. *Psychol Rev., 108*(3), 624–652.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Christophe, A., & Morton, J. (1998). Is Dutch native English? Linguistic analysis by 2-month-olds. *Developmental Science, 1*(2), 215–219.

Crain, S. (1991). Language learning in the absence of experience. *Behavioral and Brain Sciences, 4*, 597–650.

Cunnings, Ian, Patterson, Clare, & Felser, Claudia (2014). Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language, 71*, 39–56.

Dale, P. S., Dionne, G., Eley, T. C., & Plomin, R. (2000). Lexical and grammatical development: A behavioural genetic perspective. *Journal of Child Language, 27*, 619–642.

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

De Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and function words cue the acquisition of word meanings in 18-month-old infants. *Psychological Science, 30*, 319–332.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A cross-linguistic study of the relationship between grammar and lexical development. *Journal of Child Language, 32*, 759–786.

Donaldson, M., & Balfour, G. (1968). Less is more: A study of language comprehension in children. *British Journal of Psychology, 59*, 461–471.

Ehri, L. C. (1977). Do adjectives and functors interfere as much as nouns in naming pictures? *Child Development, 48*, 697–701.

Evans, G. (1980). Pronouns. *Linguistic Inquiry, 11*, 337–362.

Fenson, L., Dale, P. S., Reznick, S., Bates, E., Thal, D., Pethick, S. J., et al. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 9*(5). v-185.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*, 98–116.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology, 60*(3), 190–217.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. *Language Acquisition and Language Disorder, 44*, 97–135.

Fiengo, R. (1977). On trace theory. *Linguistic Inquiry*, 35–61.

Fodor, J. A., Bever, T., & Garrett, M. (1967). Language structure and verbal behavior. *MIT*.

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology, 26*(2), 204.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*(1), 23–45.

Goodman, M. B., & Jusczyk, P. W. (2000). The onset of sensitivity to internal syllable structure. *Papers in laboratory phonology V: Acquisition and the lexicon, 5*, 228.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav Res, 50*, 1166–1186.

Hollich, G. (2005). *SuperCoder: A program for coding preferential looking (version 1.5). [computer software]*. West Lafayette: Purdue University.

Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language, 34*, 227–249.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*, F31–F39.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language, 56*(3), 384–409.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111–116.

Kush, Dave, Lidz, Jeffrey, & Phillips, Colin (2015). Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language, 82*, 18–40.

Kush, D., Lidz, J., & Phillips, C. (2017). Looking forwards and backwards: The real-time processing of Strong and Weak Crossover. *Glossa: A Journal of General Linguistics, 2* (1), 70.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

Lakoff, G. (1968). *Pronouns and reference*. Bloomington, IN: Indiana University Linguistics Club.

Langacker, R. (1966). On pronominalization and the chain of command. (pp. 160-186). In D. Reibel, & S. Schane (Eds.), *Modern Studies in English: readings in transformational grammar*.

Lasnik, H. (1976). Remarks on coreference. *Linguist. Anal., 2*, 1–22.

Layton, T. L., & Stick, S. L. (1979). Comprehension and production of comparatives and superlatives. *Journal of Child Language, 6*, 511–527.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition, 89*(3), 295–303.

Lukyanenko, C., Conroy, A., & Lidz, J. (2014). Is she patting Katie? Constraints on pronominal reference in 30-month-olds. *Language, Learning and Development, 10*(4), 328–344.

Lust, B., Eisele, J., & Mazuka, R. (1992). The Binding Theory Module: Evidence from First Language Acquisition for Principle C. *Language, 68*(2), 333–358.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*, 339–366.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior, 2*, 346–351.

Mehler, J., & Carey, P. (1967). The role of surface and base structure in the perception of sentences. *Journal of Verbal Learning and Verbal Behavior, 6*, 335–338.

Mehler, J., & Carey, P. (1968). The interaction of veracity and syntax in the processing of sentences. *Perception & Psychophysics, 3*, 109–111.

Mehler, J., & Miller, G. (1964). Retroactive interference in the recall of simple sentences. *British Journal of Psychology, 55*, 295–301.

Momma, S., & Phillips, C. (2018). The relationship between parsing and generation. *Annual Review of Linguistics, 4*, 233–254.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005 Sep). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci., 5*(3), 263–281.

Omaki, A., & Lidz, J. (2015). Linking parser development to the acquisition of syntactic knowledge. *Language Acquisition, 22*, 158–192.

Orita, N., McKeown, R., Feldman, N., Lidz, J., & Boyd-Graber, J. (2013). *Discovering Pronouns Using Discourse Information*. Proceedings of the Cognitive Science Society.

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

R Development Core Team. (2020). R: A language and environment for statistical computing. http://r-project.org/.

Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.

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

Sag, I. (1976). *Deletion and Logical Form*. PhD Thesis. MIT.

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early understanding of subject and object Wh-questions. *Infancy, 4*(3), 423–436.

Shi, R., Legrand, C., & Brandenberger, A. (2020). Toddlers track hierarchical structure dependence. *Language Acquisition*. https://doi.org/10.1080/10489223.2020.1776010.

Smith, N. V. (1975). *Universal tendencies in the child's acquisition of phonology* (pp. 47–65). Cognitive Deficits and Retardation: Language.

Sutton, M. (2015). *Competence and performance in the Development of Prinicple C. PhD Dissertation*. University of Maryland.

Sutton, M., Fetters, M., & Lidz, J. (2012). Principle C behavior is caused by Principle C in 30 month-olds: Evidence against a transitivity bias. In *Paper presented at the 5th Generative Approaches to Language Acquisition North America (GALANA). Lawrence, KS, October 11, 2012*.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147–166.

Swingley, D., Pinto, J., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognit., 71*, 73–108.

Syrett, K., & Lidz, J. (2010). 30-month-olds use the distribution and meaning of adverbs to interpret novel adjectives. *Language Learning and Development, 6*(4), 258–282.

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

Trueswell, J., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten- path effect: Studying on-line sentence processing in young children. *Cognit., 73*, 89–134.

Wagers, M. W. (2014). Syntax in forward and in reverse: Form, memory and language processing. In A. Carnie, D. Siddiqi, & Y. Sato (Eds.), *Routledge handbook of syntax* (pp. 409–425).

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*, 2143–2152.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology, 34* (6), 1289.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer Publishing Company, Inc.

Williams, E. (1977). Discourse and Logical Form. *Linguistic Inquiry, 8*, 101–139.

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.