

An eye-tracking study examining the role of question-answer congruency in children's comprehension of *only*: A preliminary report

Lauren Covey*¹, Caitlin E. Coughlin², and Utako Minai²

¹Department of Linguistics, Montclair State University

²Department of Linguistics, University of Kansas

'Crain's puzzle' is a term that has been used to describe children's difficulty comprehending the focus operator *only* when it is in subject position (*subject-only*), showing a tendency to interpret *only* as if it preceded the verb phrase instead. While some researchers attribute children's difficulty to impoverished pragmatics in the discourse (Hackl et al., 2015), others argue that children's grammar fundamentally differs from adults' Notley et al. (2009), yielding a debate regarding whether children's misinterpretation reflects a non-adult-like linguistic representation of *only* or some computational burden on their processing of meaning. This study addresses this debate by using eye-tracking to examine whether pragmatic felicity guides children's eye-movements to incorporate the necessary information during processing on par with adults. Following Hackl et al. (2015), we experimentally manipulated whether the prompt question preceding the target sentence is pragmatically congruent or incongruent in felicitously introducing the *only*-statement in terms of which element in the sentence is focused by *only*. Emerging findings reveal that pragmatic richness in the discourse affected processing in both adults and children in a condition that was logically false. Results thus far provide support for an account which posits an important role for pragmatics.

Keywords: comprehension of focus words, preschool-age children, visual world eye-tracking

1. Introduction

In examining children's interpretation of sentences containing the word *only* in the pre-subject position (e.g., *Only Pooh ate an apple*), studies have shown that children often interpret these sentences as if *only* were placed before the verb phrase (VP) instead (e.g., *Pooh only ate an apple*) (e.g., Crain et al., 1994, 1992). Some researchers have attributed children's tendency to 'spread' the interpretation of *only* to the VP to a non-adult-like grammatical representation of *only*; that is, it has been suggested that there are fundamental differences between adult and child grammars in the instantiation of *only* (e.g., Notley et al., 2009). In contrast, other proposals suggest that children's performance on sentences containing *subject-only* relates to processing difficulty incurred by pragmatic infelicity (e.g., Hackl et al., 2015). While several studies have examined children's comprehension of *subject-only*, most of this work has used offline methods such as truth value judgments. Less is known regarding children's real-time processing of *only* (Paul et al., 2017), and thus the current study adopts the visual world eye-tracking paradigm to investigate children's online performance regarding *subject-only* to better evaluate competing proposals. We compare children and adults with respect to both the truth value judgments of sentences with *subject-only* and the real-time processing of those sentences.

1.1. Focus operators: *Only*. Focus operators express contrastive focus, indicating that some element is intended to be contrasted with a set of alternative elements (Rooth, 1992). In English, the language of interest in the current study, the focus operator *only* is used to contrast an explicit, focused element with a

*Address correspondence to coveyl@montclair.edu

set of alternatives in the discourse. For example, in the sentence *Only Pooh ate an apple*, Pooh is considered the focused element (i.e., the character who ate an apple) and is therefore contrasted with a set of alternative characters in the domain who did not eat an apple.

The syntactic context is used to indicate which constituent of the sentence is considered focused; this is specified in terms of *scope*, the linguistic domain in which the focus operator functions. In the example sentence, *only* takes scope over the subject *Pooh*. Note that *only* can also appear in other syntactic positions, including in pre-verbal position, such as in the sentence, *Pooh only ate an apple*; hence, *only* can also take scope over the whole VP. In English (and many other languages), focus operators are restricted to take scope over constituents within their syntactic c-command domain. Thus, in pre-subject position, such as in (1a), *only* c-commands the subject noun phrase (NP), while in pre-verbal position in (1b), *only* c-commands the VP.

- (1) a. [Only Pooh] ate an apple.
b. Pooh [only ate an apple].

The semantic instantiation of *only* is equally important with regard to its interpretation. Following Horn (1969), the focus operator *only* is claimed to have two independent meaning components necessary for interpretation. Adopting terminology from Horn (1969), we call the first meaning component the *presupposition*, which provides true information about the phrase associated with the focus operator; we call the second meaning component the *assertion*, which yields polarity opposite to the polarity yielded in the presupposition. For *only* sentences in positive polarity (like those given in (1)), the assertion covertly creates a negative statement regarding a set of alternatives *not* in the discourse. That is, the assertion states that a set of alternatives in the discourse, i.e., a *contrastive set*, is not associated with the focused phrase. For example, (1) above can be represented as (2) below:

- (2) Only Pooh ate an apple.
a. Presupposition: Pooh ate an apple.
b. Assertion: No one other than Pooh ate an apple.

On the basis of these two meaning components, for (2) to be true, both the presupposition (2a) and assertion (2b) must be true. With respect to the presupposition, the focused character (Pooh) must have eaten an apple; with respect to the assertion, it must be true that no other characters in the domain could have eaten an apple. The semantic representation of *only* is thus complex, requiring instantiation and evaluation of the two-faced independent meaning components. During comprehension of sentences containing focus operators, a series of complex representations must be constructed, which is not the case for sentences without focus operators.

One important issue in the acquisition literature regards how children comprehend *only*. As language learners, children need to learn both the syntactic and semantic linguistic properties of *only* in order to interpret sentences containing *only* in an adult-like manner. Investigating whether and to what extent children may be aware of these linguistic properties creates interesting testing grounds, through which researchers can observe important developmental aspects of child syntax and semantics. Previous studies examining children's interpretation of *only* have reported that young children up to the age of 6+ exhibit variability in interpreting sentences containing *only*; in the following sections, this literature is briefly reviewed.

1.2. Children's interpretation of *only*: A tendency to 'spread' from subject to VP. In the first studies to report children's misinterpretation of *only*, Crain and colleagues (Crain et al., 1992, 1994) found that when *only* preceded the subject of the sentence (hence referred to as *subject-only*), such as in (3a), children tended to incorrectly interpret the sentence as if the focus operator took scope over the VP, as in (3b).

- (3) a. Only the cat is holding a flag.
b. The cat is only holding a flag.

Using a truth value judgment task, Crain and colleagues found that children tended to incorrectly accept sentences like (3a) when viewing a picture in which a cat was holding a flag and nothing else (which

verifies (3b)), even though another character in the discourse context was also holding a flag (which crucially falsifies (3a)). In a comparison condition utilizing the same target sentence (3a), a picture shown to the children critically differed in that the cat was holding *two* objects (e.g., a flag and a balloon), logically falsifying the sentence. In this condition, children tended to incorrectly reject sentences like (3a), even though no other character in the discourse context was holding a flag. When asked to explain their rejection of the sentence, children provided non-adult-like justifications by describing the fact that the mentioned character was holding two objects. This pattern reveals that, overall, children incorrectly associate the subject-*only* with the VP; in the current paper, we will refer to this type of error pattern as a 'spreading error'.

On the other hand, it has not been consistently shown that children incorrectly associate VP-*only* with the subject noun phrase. The asymmetry between children's difficulty comprehending subject-*only* and their ease in comprehending VP-*only* has been labeled "Crain's puzzle" (Hackl et al., 2015; Sugawara, 2016). Specifically, children have been shown to have an adult-like interpretation of VP-*only*, yet incorrectly interpret subject-*only* sentences as if the focus operator took scope over constituents in the VP (i.e., interpreting 3a as if it were 3b). A variety of studies to date have observed this asymmetry in English (Philip & Lynch, 2000; Paterson et al., 2003), German (Müller et al., 2011), Mandarin (Notley et al., 2009; Zhou & Crain, 2010), and Japanese (Endo, 2004; Sano, 2011).

Several hypotheses¹ have attempted to explain children's interpretation of subject-*only* as taking scope over constituents within the VP; in the current paper, we will focus on two hypotheses that lead to alternating predictions. One proposal states that children's syntactic representation of *only* differs from that of adults; this hypothesis argues that there are fundamental differences between adult and child grammars in the instantiation of *only*. On the other hand, a second account proposes that children's asymmetrical performance on subject- and VP-*only* is rooted in processing issues related to pragmatic (in)felicity. In the following section, these two proposals regarding Crain's puzzle are briefly reviewed.

2. Literature review

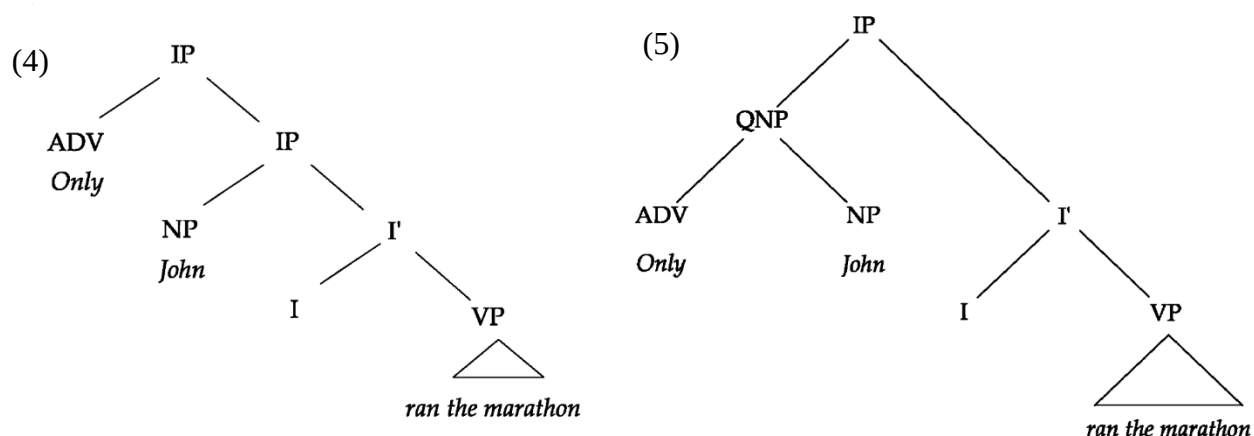
2.1. Sentential analysis hypothesis. In their study first reporting children's asymmetrical misinterpretation of *only*, Crain et al. (1994) hypothesized that the difference between children's interpretation of subject- and VP-*only* was driven by differences in the grammar of adults and children, such that children's grammar *only* allowed for a VP-oriented analysis of *only*. This hypothesis was further specified in two studies testing Mandarin-speaking children by Notley et al. (2009) and Zhou & Crain (2010). Examining performance using the Mandarin focus operator *zhiyou* (which behaves similarly to *only*), Notley et al. (2009) found that Mandarin-speaking children exhibited a similar asymmetry in their comprehension of *zhiyou*, with adults interpreting subject-*zhiyou* sentences using a subject scope analysis and children largely interpreting subject-*zhiyou* as taking scope over the VP. Crucially, children's poor performance on sentences containing subject-*zhiyou* became increasingly more adult-like in a follow-up experiment. In the subsequent experiment, a bias manipulation was introduced to test whether children who displayed a strong VP-scope interpretation could exhibit an adult-like subject interpretation when the experiment biased them toward a subject interpretation. The bias manipulation was operationalized such that at the beginning of each trial, the child (and a storytelling puppet) was asked who (i.e., which character) they thought would do something in the story, prior to the action itself (e.g., *Who do you think can jump over the two things?*). At the end of the story, the puppet produced the target sentence in response to the question, *Who did something?* rather than to the more broad question, *What happened?*. These two experimental changes served to highlight the subject of each story in order to encourage a subject interpretation of the focus operator. Children's performance on the subject-*zhiyou* conditions markedly increased from 10% accuracy to 76.2% in this follow-up experiment. Moreover, in the subject-*zhiyou* false condition, children correctly rejected the sentences by providing an adult-like subject scope responses 61.9% of the time. Note that while children's responses became increasingly more adult-like in this experiment, the VP-scope interpretation was not ruled out entirely, given that some children consistently provided VP-scope responses throughout (e.g., 38.1% of justifications remained VP-oriented).

In summary, Notley et al. (2009) provided crosslinguistic evidence from Mandarin-speaking children

¹See also 'No Contrast Computation' (Paterson et al., 2003) and 'Focus-Default' (Müller et al., 2011).

that the focus operator *zhiyou* in pre-subject position is often analyzed as taking scope over the VP. These same Mandarin-speaking children seemed to override their VP-scope interpretation when subject-biasing questions were included in the experimental stories. To explain children's dominant VP-scope interpretation, Notley et al. (2009) hypothesized that children process subject focus operators as sentential adverbs. Specifically, if the focus operator were to reside in the adverbial position in the syntactic tree, it would c-command the subject NP and the VP, allowing the focus operator to take scope over both, on par with the sentential adverb (see (4) below). Thus, under this proposal, children's interpretation of *zhiyou/only* taking scope over the VP is possible due to the sentential adverb position which c-commands the VP.

In the adult representation, the c-commanding domain of subject-*only* is limited to the subject NP (see (5) below); in children's representation, however, the pre-subject focus operator c-commands the VP as well as the NP (4). Hence, children whose syntactic representation allows the pre-subject focus operator to c-command the VP robustly interpret the pre-subject focus in association with a VP element. It should be highlighted that the Sentential Analysis Hypothesis attributes children's non-adult-like comprehension of the pre-subject focus to their non-adult-like linguistic representation; children's grammar is fundamentally different from an adult's in the grammatical representation of the focus operator.

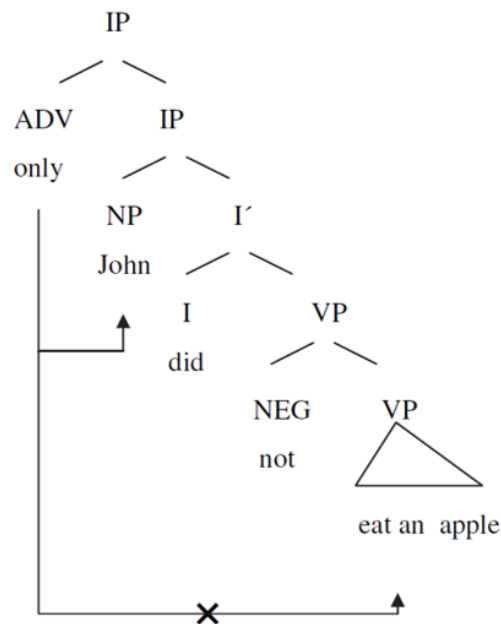


Zhou & Crain (2010) strengthened the claim of the Sentential Analysis Hypothesis by evidencing that children's tendency to wrongly associate a pre-subject focus with the VP is lowered when the pre-subject focus no longer c-commands the VP. They manipulated the syntactic structure of test sentences containing pre-subject focus by including a negation word intervening between the NP and VP (e.g., *Zhiyou Yuehan meiyou chi pingguo*; 'Only John did not eat an apple'). In this linguistic environment, the association between the focus operator and the VP is blocked because both the focus operator and negation word have the same structural role in the syntactic representation (i.e., in adverbial position); this is known as a Relativized Minimality effect (Rizzi, 1990, 2001). Under Relativized Minimality, local syntactic relations (e.g., between *zhiyou* and *chi pingguo* 'eat an apple') are blocked due to an intervening element (e.g., *meiyou* 'not'), which structurally does not allow for a c-commanding relationship between the focus operator and the VP. Hence, if children interpret the focus operator as an adverb, as predicted by the Sentential Analysis Hypothesis, then the negation word *meiyou* should block the relationship between *zhiyou* and the VP, resulting in a subject-scope interpretation. In a truth value judgment task, Zhou & Crain (2010) found that Mandarin-speaking children correctly rejected subject-*zhiyou* false sentences with negation 100% of the time, providing an adult-like falsification justification relating to a subject-scope analysis.

When negation is present, creating a Relativized Minimality effect, children's tendency to associate the pre-subject focus with the VP element is suppressed, as the c-commanding which is required to associate the focus with the VP element is blocked. This blocking effect is demonstrated in (6) below (from Notley et al., 2009, example 39b), where *not* blocks the relationship between *only* and the VP, *eat an apple*. As discussed by Notley et al. (2009), this argument has broad implications that child and adult grammars differ from

one another in regards to how focus is realized syntactically. Zhou & Crain (2010) argue that there is a developmental stage that children go through at which they instantiate subject-*zhiyou* as an adverb taking scope over both the NP and the VP.

(6)



In contrast to these findings, there has been some evidence challenging the Sentential Analysis Hypothesis in studies which have shown that children's interpretation of subject-*only* is highly adult-like. For example, Goro et al. (2005) examined English-speaking children's interpretation of sentences containing the disjunction operator *or*, such as in *Only Bunny Rabbit ate a carrot or a green pepper*. In Goro et al. (2006), target sentences contained the conjunction operator *and*, for example, *Only Aladdin opened the blue box and the black box*. These sentences provide insight into whether children correctly compute the assertion component of *only*. Recall that the assertion component covertly creates a negative environment (Horn, 1969). Because logical connectives (e.g., *and/or*) interact with negative elements, computing the assertion component of the sentence becomes difficult, as shown in (7-8) below.

- (7) Only Bunny Rabbit ate a carrot or green pepper.
 a. Presupposition: Bunny Rabbit ate a carrot or a green pepper.
 b. Assertion: No one else ate a carrot and no one else ate a green pepper.
- (8) Only Aladdin opened the blue box and the black box.
 a. Presupposition: Aladdin opened the blue box and the black box.
 b. Assertion: No one else opened the blue box or the black box (i.e., "not both").

These examples provide a good test case of whether children are able to correctly interpret the logical connective by computing a correct presupposition and assertion of the focus operator. Results of both experiments indicated that children indeed interpreted the target subject-*only* sentences in an adult-like way. In Goro et al. (2005), English-speaking children correctly accepted true sentences containing the logical operator *or* 93% of the time and correctly rejected false sentences 90% of the time. In Goro et al. (2006), English-speaking children correctly accepted true sentences containing the logical operator *and* 95% of the time and correctly rejected false sentences 85% of the time. Children's acceptance rates were largely adult-like, and indicate that in the presence of a logical connective, children correctly interpret subject-*only*.

These studies provide counterevidence to the Sentential Analysis Hypothesis because they show that children can perform at adult-like levels in their interpretation of subject-*only* sentences which do not have

a Relativized Minimality effect (Rizzi, 1990, 2001). While results from Goro et al. (2005, 2006) are similar to those of Notley et al. (2009) and Zhou & Crain (2010) in that they show children's adult-like performance on subject-*only*, what is crucially different about stimuli from Goro and colleagues is that their sentences containing a logical connective do not involve a Relativized Minimality effect (i.e., a blocked association between *only* and the VP due to a third linguistic element). Under the Sentential Analysis Hypothesis, it remains unclear why children stop associating the pre-subject *only* to VP elements, even when there is no syntactic representation that blocks the c-commanding between the pre-subject *only* and the VP.²

2.2. Pragmatic felicity hypothesis. Let us now turn to the second hypothesis, which attempts to account for Crain's puzzle from a pragmatic perspective. The *Pragmatic Felicity Hypothesis* does not hypothesize a deficit or misrepresentation in children's syntactic knowledge, and subsequently does not explore Relativized Minimality effects in children's interpretation. Instead, it accounts for children's performance on subject-*only* at the level of pragmatic felicity and online computation. A recent study conducted by Hackl et al. (2015) (see also Sugawara, 2016) manipulated pragmatic felicity via the Question-Answer Congruency (QAC) principle (Paul, 1880; Rooth, 1985, 1992). This principle states that answers to *wh*-questions must contain a constituent that corresponds to the focused element within the question. In (9) below, an example from Sugawara (2016) illustrates the QAC principle.

- (9) a. Q: Who is holding a flag?
 b. A1: THE CAT_F³ is holding a flag.
 c. A2: *The cat is holding A FLAG_F.

The question (9a) pragmatically requires an answer about the subject of the sentence, and the answer (9b) is felicitous because it complies with the QAC principle given that the subject is focused. The answer in (9c) is incongruent because the focused element is in object position, despite the fact that the question (9a) requires information about the subject position.

The QAC principle is of particular relevance to the investigation of children's interpretation of *only*. Hackl et al. (2015) pointed out that previous studies employed target sentences which did not pragmatically follow from the questions asked. For example, the truth value judgment tasks (i.e., sentence-picture matching) typically involved a puppet character that interacted with the children and the experimenter. After observing a story ending with a scene or picture, the experimenter asked the puppet, *What happened?*. The puppet would then utter the target subject-*only* sentence such as (3a), *Only the cat is holding a flag*. Crucially, from a QAC perspective, the target subject-*only* sentence is not fully congruent with the question *What happened?* because the focused element in the question (e.g., *what*) does not directly correspond to the answer provided by the puppet, which relates to the character. It is important to note that the answer is not necessarily fully infelicitous, because there are sub-questions that follow directly from the asked question *What happened?* which result in an ultimately congruent question-answer pair (Sugawara, 2016). For illustration, see (10) below:

- (10) a. What happened?
 b. Who is holding the flag?
 c. What is the cat holding?

The sub-questions (10b) and (10c) are both relevant to the question under discussion (10a). Note that one possible pragmatically felicitous answer to the sub-question (10b) is the subject-*only* sentence shown above in (3a), *Only the cat is holding a flag*. Because the answer (3a) satisfies a sub-question of the overall question (10a) *What happened?*, the QAC is indirectly satisfied.

²Goro and colleagues did not provide any discussion regarding why children's comprehension of the sentences with subject-*only* became highly adult-like in their studies. It should be noted that what was unique about their test sentences was the inclusion of the logical connective in the VP, creating a conjoined nominal. This implies, although not directly examined or discussed, that the presence of the logical connectives may have played a role in inviting the adult-like comprehension from children. Whereas it remains an open question why the children in their study robustly avoided the 'spreading' error, we take their findings as counterevidence to the Sentential Analysis Hypothesis.

³Here, F is used to indicate a focused element.

Hackl et al. (2015) and Sugawara (2016) hypothesize that it is this construction of sub-questions (in an attempt to satisfy the QAC principle) which results in children's asymmetrical performance on subject- and VP-*only*: children may not be sensitive to question-answer pairs which require resolution via a sub-question, and instead, they are better able to compute meaning when an answer is directly congruent with its question. Such a proposal is able to account for children's non-adult-like performance on subject-*only* in previous studies, because in these experiments the target sentence was preceded by the prompt question *What happened?* (e.g., Crain et al., 1994). Hackl et al. (2015) suggest that previous experiments which employed questions that did not directly satisfy the QAC principle may have underestimated children's ability to interpret subject-*only*. In addition, this proposal is able to explain why children typically do not display difficulty with sentences containing VP-*only* in these experiments: both the sub-question in (10c) and the overall question (10a) are 'what' questions, and thus it is more straightforward to construct the sub-question (10c) as compared to the 'who' question in (10b), which is not a sub-question related to the prompt question *What happened?*. Hence, constructing sub-questions with respect to the question under discussion is more challenging for subject-*only* sentences, in which the sub-question (10b) does not necessarily directly follow. In summary, this hypothesis predicts that children's comprehension is intact at the grammatical level, but that when pragmatics are impoverished in the discourse, errors are expected to emerge. Specifically, when the QAC principle is directly rather than indirectly satisfied, children's performance on subject-*only* should become more adult-like.

Hackl et al. (2015) (see also Sugawara, 2016) empirically tested this proposal by including trials in which the context question is pragmatically congruent with the target sentence (e.g., *Who is holding the flag?*) as well as trials which contained incongruent questions (e.g., *What happened?*). Data from English-speaking children indicated that there were significantly more adult-like interpretations following trials with pragmatically felicitous questions as compared to infelicitous ones.⁴ In fact, children provided accurate responses to over 70% of congruent subject-*only* trials. The researchers also included VP-*only* target sentences that manipulated QAC. Interestingly, when the question was incongruent with a VP-oriented *only* (e.g., *Who is holding the flag?*), children incorrectly interpreted the sentence as taking scope over the subject, providing a non-adult-like response 69% of the time. This second set of results further highlights the importance of QAC, given that few previous studies had shown a non-adult-like pattern of performance for VP-*only* sentences. This suggests that children's comprehension of subject- and VP-*only* is likely affected by children's sensitivity to pragmatic felicity, rather than to a difference in the syntactic representations.

In a similar experiment, Sugawara (2016) tested adult native English speakers in a parallel study in which the prompt questions were either congruent or incongruent with the QAC principle. Crucially, adults' accuracy rates were modulated in the same way as children's: on subject-*only* sentences, adults were highly accurate in responding to the QAC congruent sentences (94%), but accuracy decreased to 73% for the incongruent subject-*only* condition. This pattern extended to the VP-*only* condition, such that the congruent accuracy rate of 90% decreased to 66% in the incongruent condition. These findings suggest that adults similarly have difficulty comprehending sentences containing *only* when pragmatics in the discourse are impoverished. This is further supported in the reaction time (RT) data: an asymmetry between subject- and VP-*only* emerged, with adults showing increased RTs for subject-*only* across experiments as compared to the VP-*only* counterparts. This pattern was taken to suggest subject-*only* sentences (in response to the QAC-incongruent question, *What happened?*) incur a processing cost compared to VP-*only* sentences.

Overall, these studies provide evidence that both children and adults are sensitive to the Question-Answer Congruency principle, such that congruency between a question and its answer influences whether interpretations are target-like or not. To further investigate this proposal, the present study builds on Sugawara's (2016) reaction time data in order to directly examine the processing sentences containing subject-*only*. This study is one of the first to examine children's processing of *only* using an online methodology; the goals and research questions of the present study are outlined in the following section.

⁴In their follow-up experiment, Notley et al. (2009) also utilized a pragmatic manipulation, and similarly reported that children show increasingly adult-like interpretations for sentences containing subject-*only* (*zhijou*).

3. Current study

As reviewed above, most studies examining Crain's puzzle thus far have utilized offline methods (e.g., truth value judgments) to assess children's interpretation of sentences containing *only*, and less is known regarding the real-time processing aspect of children's *only* comprehension. One study thus far has reported a discrepancy between the offline comprehension measures and online eye-movement measures for children's *only* comprehension (Zhou et al., 2012). Their study targeted the use of prosodic stress in the resolving ambiguity of the scope of *zhiyou* ('only'). Although children often provided non-adult-like end-of-sentence responses in the actual meaning judgment, their eye movement patterns closely matched those of adults, albeit at a delay. This suggests that relying primarily on offline judgements may misestimate children's meaning processing, given the asymmetry in children's adult-like eye fixations and non-adult-like judgments, and that other factors (such as pragmatic congruency) may affect their interpretations. Thus, the discrepancy between children's offline meaning judgment and online eye movements during meaning processing underscores the importance of the addition of eye movement measures to the judgment tasks utilized thus far.

Based on these discussions above, the current study adopts the visual world eye-tracking paradigm to investigate children's online performance with respect to their offline judgments regarding subject-*only*. Note that Hackl et al. (2015) revealed an effect of QAC manipulation, in which children's comprehension of subject-*only* became more adult-like when the *only*-sentence was introduced by a *wh*-question congruent to the subject focus. The current study builds upon their findings, adding an eye-movement measure to further examine whether the QAC manipulation effect may guide children's eye movements to incorporate the necessary information into the meaning processing on par with adults, as well as their truth-value judgments presented as the end-state of meaning processing. Following Hackl et al. (2015), we experimentally manipulated whether or not the prompt question preceding the target sentence was congruent or incongruent with the QAC. This allowed us to examine to what extent children's comprehension of subject-*only* may be adult-like, as well as to revisit a discussion regarding whether children have processing issues related to pragmatic infelicity, such that adding prompt questions which are congruent with the QAC improves children's interpretation (Hackl et al., 2015; Sugawara, 2016), or whether children ultimately consider subject-*only* as taking scope over the VP regardless of the pragmatic richness in the discourse (Notley et al., 2009; Zhou & Crain, 2010).

We designed a sentence-story verification task based on the conventional paradigm of the Truth Value Judgment task (TVJ, Crain & Thornton, 1998). In the TVJ story-setting, an English-learning character (an Alien) is introduced to the child and views stories with the child. At the end of each story, the Alien described what he thinks happened in the story, which serves as a test sentence. The child's task is to decide whether the character correctly described the story, which in turn reflects the child's own truth value judgment of the test sentence based on the story context. In tandem with this sentence-story verification task, we administered the visual-world eye-tracking paradigm, which allowed us to further examine the real-time aspects of children's sentence-story verification procedure, i.e., the processing of their sentence meaning comprehension, as well as the ultimate outcome of their meaning comprehension. Visual world eye-tracking is an experimental method in which eye movements are tracked as participants listen to utterances while viewing a relevant scene. Adults (Tanenhaus et al., 1995) and children (Trueswell et al., 1999) launch eye movements fixating on appropriate images or objects as a sentence unfolds and fixations to relevant objects closely follow reference to abstract linguistic representations. Eye-tracking provides insights into language comprehension and the precise time-course (at the millisecond level) of this complex process.

3.1. Participants. Sixteen English-speaking children (mean age 5;04, range 5;0–5;10, 8 males) were recruited in the surrounding community. From this group, data from five children was excluded from further analysis. In the case of three of these participants, extreme data loss occurred during the experiment due to excessive blinks or technical difficulty in calibration. The fourth subject was excluded because he/she asked to end the experiment early; data from one participant was excluded due to inaccurate responses to all experimental stimuli. After excluding these five children, the total number of participants was 11 (mean age 5;04, range 5;0–5;10, 4 males). Each child was randomly assigned to the Control or Experimental group; after exclusion of the five children, this resulted in 5 participants in the Control group (mean age 5;05) and 6 participants in the Experimental group (mean age 5;04). While the goal of the study was to test monolin-

gual English-speaking children, it was the case that some of the children tested also knew other languages: one child in the Control group was bilingual in Mandarin Chinese and English, and spoke Mandarin in the home; one child in the Experimental group was familiar with Spanish, Navajo, and English, speaking all three in the home. In an effort to examine as much data as possible, data from these bilingual children were included in the analyses. All children received a small token for participating, and their parents were paid \$10 per visit.

Twenty-three adult native speakers of English were recruited from the University of Kansas. Adults were randomly assigned to the Control or Experimental group. The Control group consisted of 12 participants (5 males), with a mean age of 21.8 (range 18-34). The Experimental group consisted of 11 participants (4 males), with a mean age of 20.8 (range 19-23). All adult participants either received course credit or \$5 for participating.

3.2. Materials. The target stimuli consisted of sentences with subject-only, as in (11).

(11) Only the Genie got a banana.

Three target conditions were created by manipulating the truth conditions of each sentence, based on predicted comprehension patterns of both adults and children. The three subject-*only* conditions are displayed in Table 1, with examples of context sentences, which were presented prior to the target sentence.

Condition	Context	Target Sentence (subject- <i>only</i>)
1. Adult <i>true</i> / Child <i>false</i>	Oh, the Fairy got a cookie and a strawberry. Oh, the Genie got a banana and a <u>cookie</u> . Oh, the Wizard got a strawberry.	Only the Genie got a banana .
2. Adult <i>false</i> / Child <i>true</i>	Oh, the Wizard got a plate and a bell. Oh, the Fairy got a crayon . Oh, the Genie got bell and a <u>crayon</u> .	Only the Fairy got a crayon .
3. Adult <i>false</i> / Child <i>false</i>	Oh, the Fairy got a TV and a pretzel. Oh, the Genie got a TV and a box. Oh, the Wizard got a pretzel and a <u>box</u> .	Only the Wizard got a pretzel .

Table 1: Target experimental conditions.

In the adult-true condition (Condition 1), the outcome of the sentence/preceding discourse is that the Genie is the only character who received a banana, which logically makes the target sentence true; thus, a 'true' interpretation is the predicted adult-like response for such a trial. Crucially, this condition may invite a false interpretation from children if they 'spread' the subject-*only* to the VP and interpret the sentence as if there were a VP-*only*, because the mentioned character, Genie, received both a banana and a cookie. In the adult-false/child-true condition (Condition 2), the outcome of the sentence and preceding discourse is that the Fairy is not the only character who received a crayon, rendering the target sentence logically false and resulting in an adult-like 'false' interpretation. In this condition, the target character (Fairy) received only the target object (a crayon), and thus a 'spreading' interpretation from children would result in acceptance of the sentence, given that under a VP-*only* analysis, it is true that the only object the Fairy received is a crayon. Condition 3 is also considered an adult-false condition, because the outcome of the sentence and preceding discourse is that a character other than the Wizard also got a pretzel, rendering the sentence logically false. In addition, if children spread *only* and interpret it as VP-*only*, children are also predicted to reject the sentence given that, in the discourse context, the Wizard received a second object besides the pretzel. To distinguish between the adult and child false responses, truth value judgments, and justification for false responses (which indicates whether the child participant's falsification is indeed adult-like or reflects a 'spreading' error for Condition 3) were analyzed.

Four trials per condition ($n=12$ total) were included. Ten VP-*only* trials were also included in the experiment as fillers (e.g., *The Fairy only got a pen*). Six of the VP-*only* trials were logically true based on the

preceding context and the other four *VP-only* trials were logically false. Finally, 11 trials in which the characters each received only one object (e.g., *The Genie got a sweater*) were also included as a control condition intended to ensure that children correctly comprehended sentences without *only* (5 trials were logically true and 6 were logically false). This resulted in a total of 33 trials divided equally across three blocks ($n=11$ per block). The first block contained all control trials, and Blocks 2 and 3 were pseudorandomized with both target (subject-*only*) and filler (*VP-only*) trials. This allowed children to become familiarized with the experiment format on the control trials in Block 1 before being exposed to the target trials in Blocks 2 and 3.

In addition to the within-subjects factor Condition (Condition 1 vs. Condition 2 vs. Condition 3), this study adopted a mixed design to test the between-subjects factor QAC congruency (Control group vs. Experimental group). Following Hackl et al. (2015), two types of context questions (the prompt question eliciting the test sentence utterance by Alien, which was presented between the Context and Test sentence presentation) were included, shown in (12).

- (12) a. Alien, who got a banana? – (QAC-congruent, Experimental group)
 b. Alien, what happened? – (QAC-incongruent, Control group)

In (12a), the focused element within the question (e.g., *who*) directly corresponds to the answer provided by the alien in the subject-*only* (11). In (12b), the focused element in the question (e.g., *what*) does not directly correspond with the subject-*only* answer provided by the alien and thus the target sentence (12b) is considered incongruent. QAC was manipulated in a between-subjects design, such that participants either encountered target sentences that were always preceded by the broad question in (12b) or with QAC-congruent questions such as in (12a).⁵

For the visual materials (i.e., pictures), the three characters mentioned in Table 1 were repeated across all trials (Fairy, Genie, Wizard). All objects selected by the characters were imageable⁶ and normed via Age of Acquisition (AOA) database such that children aged 5 and older would be familiar with them (Kuperman et al., 2012). Figure 1 provides an example of a visual display for a target trial (see Table 1, Condition 1). Characters' objects always appeared to the right of the character. The black boxes shown in the figure surrounding the characters and objects demonstrate the Areas of Interest (AOI) for data analysis but did not appear on the screen during presentation.

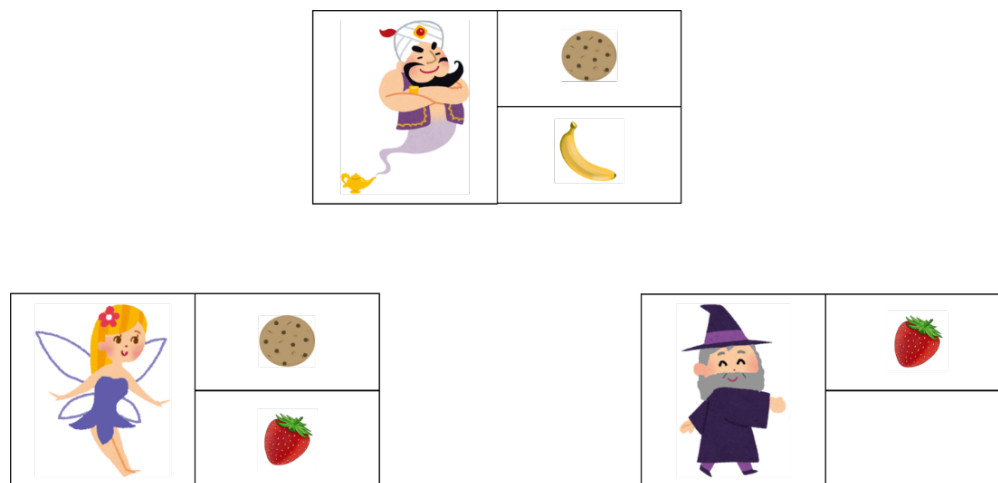


Figure 1: Example visual display of trial (with segmented AOIs).

⁵To ensure that the experiment was an appropriate length for children to participate comfortably, each condition contained a relatively low number of trials (as summarized above). Therefore, we manipulated QAC as in a between-subjects design to ensure that there were no exposure/practice effects from participants being exposed to QAC-congruent and QAC-incongruent contexts within the same short experiment.

⁶Images were selected from the following open source websites: <https://openclipart.org/>, <http://www.mycutegraphics.com/graphics/school-images.html>, <http://www.irasutoya.com/>, <http://girlysozai.com/>

All sentences were presented auditorily, including discourse contexts, context questions, and target sentences. Audio was recorded prior to testing and normalized for amplitude using Praat (Boersma, 2001). Two female native English speakers read the target materials at a slow but natural speech rate. The first speaker served as the narrator of the story and read the three sentence discourse contexts and uttered the context question. The second speaker uttered the target stimulus sentences, serving as the 'alien' (see Procedures below). For the eye-tracking analyses, target sentences were segmented into four time windows (TWs), shown in (13) below.

(13) [TW₁ Only][TW₂ the Genie][TW₃ got][TW₄ a banana].

Target sentences differed only in the mentioned character and mentioned object. Because the target sentences were recorded at a slow-but-natural rate, each trial differed slightly in the length of time from the beginning of the trial until the final noun was spoken. Across conditions, the mean duration of Time Window 1 (e.g., *only*) was 329.4 ms (SD 30.4); Time Window 2 (e.g., *the Genie*) was 574.5 ms (SD 32.3); Time Window 3 (e.g., *got*) was 275.6 ms (SD 69.8); and Time Window 4 (e.g., *a banana*) was 543.4 ms (SD 99.2). Tables 2-4 demonstrate mean durations per critical region for each condition of separately.

	TW 1 (<i>Only</i>)	TW 2 (<i>the</i> [noun])	TW 3 (<i>got</i>)	TW 4 (<i>a</i> [noun])
Mean	328.8	577.5	273.7	540.3
SD	11.4	21.3	77.7	53.9

Table 2: Mean length of critical regions for Condition 1 (ms).

	TW 1 (<i>Only</i>)	TW 2 (<i>the</i> [noun])	TW 3 (<i>got</i>)	TW 4 (<i>a</i> [noun])
Mean	310.8	568.3	276.6	532.6
SD	14.4	29.2	76.3	108.5

Table 3: Mean length of critical regions for Condition 2 (ms).

	TW 1 (<i>Only</i>)	TW 2 (<i>the</i> [noun])	TW 3 (<i>got</i>)	TW 4 (<i>a</i> [noun])
Mean	348.5	577.7	276.6	557.5
SD	45.9	49.4	77.6	144.9

Table 4: Mean length of critical regions for Condition 3 (ms).

3.3. Procedures. At the beginning of the experiment, participants were asked if they would like to play a game with an alien who was still learning English. The participant was asked to help the alien by telling him when he said something right and correcting him when he said something wrong. This introduction set up a context in which children were felicitously asked to judge the truth of Alien's speech serving as test sentences. In the experiment, the participant and the alien watched a series of magic shows on the computer screen involving the three characters (Genie, Fairy, and Wizard). Each trial began with the discourse context in which the participant heard the narrator speak a series of three sentences, describing what each character in the discourse context received (i.e., used their magic powers to 'get') during a given trial. Each sentence was paired with an animation on the computer screen, such that the objects being described appeared in turn next to the characters that received them. Following the context, the narrator asked the alien the context question. For participants in the Control condition, the context question was always, *Alien, what happened?* (12b). For participants in the Experimental condition, the context question was congruent with the target sentence (e.g., in (12a), *Alien, who got a banana?*). Following the context question, the alien described something that happened in the story by uttering the target sentence.

Eye-movements were tracked using an EyeLink 1000 system (SR Research), using the experimental software Experiment Builder (SR Research). Prior to the experiment, the participant's eye-movements were calibrated using the 9-point calibration option, which took under five minutes. Each trial began with the auditory presentation of the context sentences, simultaneously presented with visual animations in which the objects appeared on the screen. Upon conclusion of the context portion, a small picture of the alien appeared in the center of the display, serving as a fixation point. The subject was asked to look at the alien, and when the participant's gaze was fixated on the picture for 1 second, the audio presentation of the question and test sentence began. Following the utterance of the target sentence, the experimenter asked the participant, "Was the Alien right or wrong?". Participants responded orally, and responses were recorded via a button press on the keyboard by the experimenter. If participants responded that the Alien uttered something incorrect, the experimenter asked for follow-up information, asking "Why was he wrong?". These justifications were manually recorded by the second experimenter.

Upon arrival at the laboratory, adult participants provided their informed consent. Parents of child participants read and signed a consent form, and minor assent was also obtained. Parents received \$10 for their participation, and children received a small token for his/her participation. Adult participants received course credit or \$5 for participating. The total length of the visit for adult participants was about 20 minutes. The total length of the laboratory visit for children was about 45 minutes, including breaks.

3.4. Data analysis. Behavioral responses, or truth value judgments of the test sentences, were coded as correct (i.e., 'right' responses for logically true items and 'wrong' responses for logically false items) or incorrect (i.e., 'wrong' for logically true items and 'right' for logically false items). 'Wrong' responses were also coded for the reason for rejection (e.g., as 'adult-like' for logical rejections and 'non-adult-like' for illogical rejections).

For eye-movement data, eye movements were time-locked to the auditory segments (1-4) in the target sentences. This preliminary analysis examines the average number of fixations to the relevant AOIs within a given segment. Averages were calculated for each participant individually. For each trial, fixations were averaged at up to four relevant AOIs: (i) mentioned character, (ii) mentioned object, (iii) competitor character, and (iv) competitor object. In Table 5, Table 1 is repeated, with the last column showing the AOIs. The mentioned character and object correspond to the target sentence utterance. The competitor character (iii) corresponds to the other character in the discourse who also obtained the mentioned object (e.g., in Condition 2, the Genie also gets a crayon). The competitor object corresponds to the other object obtained by the mentioned character (e.g., in Condition 1, the Genie also gets a cookie). We will discuss, in the following section, how eye fixations to these AOIs (particularly (iii) and (iv) in Condition 3) are interpreted with respect to predicted eye movement patterns.

Condition	Context	Target Sentence (subject-only)	AOI
1. Adult <i>true</i> / Child <i>false</i>	Oh, the Fairy got a cookie and a strawberry. Oh, the Genie got a banana and a <u>cookie</u> . Oh, the Wizard got a strawberry.	Only the Genie got a banana .	(i) Genie (ii) Genie's banana (iii) N/A (iv) Genie's cookie
2. Adult <i>false</i> / Child <i>true</i>	Oh, the Wizard got a plate and a bell. Oh, the Fairy got a crayon . Oh, the Genie got bell and a <u>crayon</u> .	Only the Fairy got a crayon .	(i) Fairy (ii) Fairy's crayon (iii) Genie (iv) N/A
3. Adult <i>false</i> / Child <i>false</i>	Oh, the Fairy got a TV and a pretzel. Oh, the Genie got a TV and a box. Oh, the Wizard got a pretzel and a <u>box</u> .	Only the Wizard got a pretzel .	(i) Wizard (ii) Wizard's pretzel (iii) Fairy (iv) Wizard's box

Table 5: Example of target experimental conditions with AOIs.

The eye-movement data were analyzed using linear mixed-effects models with the *lme4* package (Bates et al., 2014) in the R programming environment, with *p*-values calculated using the *lmerTest* package (Kuznetsova et al., 2015). For this preliminary analysis, we ran separate models for each time window in the target sentence (see (13) above for segmentation of target sentence), which resulted in four total models per Group.

The dependent variable in each model was the mean fixations at a given AOI, for a given segment. Model fitting began by including the fixed factors Group (Control, Experimental), AOI, and the interaction of Group×AOI; Subject and Item were included as random intercepts. The model was then progressively backwards-fit by hand, such that interactions and fixed effects which did not explain a significant portion of the variance were removed using log-likelihood ratio values. This allowed us to arrive at the simplest model that best fit the data. We consider $p < 0.05$ to be significant. In the following section, specific predictions regarding AOI for Condition 3 are described in detail.

3.5. Predictions. Given the scope of this preliminary report, the current paper will discuss findings for only Condition 3, the *Adult-false/Child-false* condition containing two potential falsifying competitors. This condition is of particular interest because it can be used to tease apart eye-movement patterns related to adult-like (logical) falsification and those related to child-like (incorrect) falsification driven by the ‘spreading’ of the subject-*only*, both of which would independently (and maybe distinctively) lead comprehenders to the same ‘false’ response as off-line measure of the meaning judgment. The two variables of interest are meaning comprehension, as assessed via the truth-value judgment task, and eye-movement patterns, as measured by mean fixations to the relevant AOIs during the auditory presentation of the test sentence. For meaning comprehension, it is predicted that adults will have very high accuracy comprehending the target subject-*only* sentences. Note that as in Sugawara (2016), it is possible that we will observe increased accuracy for adult participants in the QAC-experimental condition as compared to the control condition. For children, it is predicted following Hackl et al. (2015) that there will be a significant effect of pragmatic felicity, such that children in the experimental condition will show higher overall accuracy than children in the control condition. In addition, children’s non-adult-like justifications in the truth-value judgment task will provide insights about their interpretation of subject-*only*, and it is specifically predicted that non-adult-like responses will primarily result from ‘spreading’ errors, in which children interpret subject-*only* as taking scope over the VP.

(14) Only the Wizard got a pretzel.

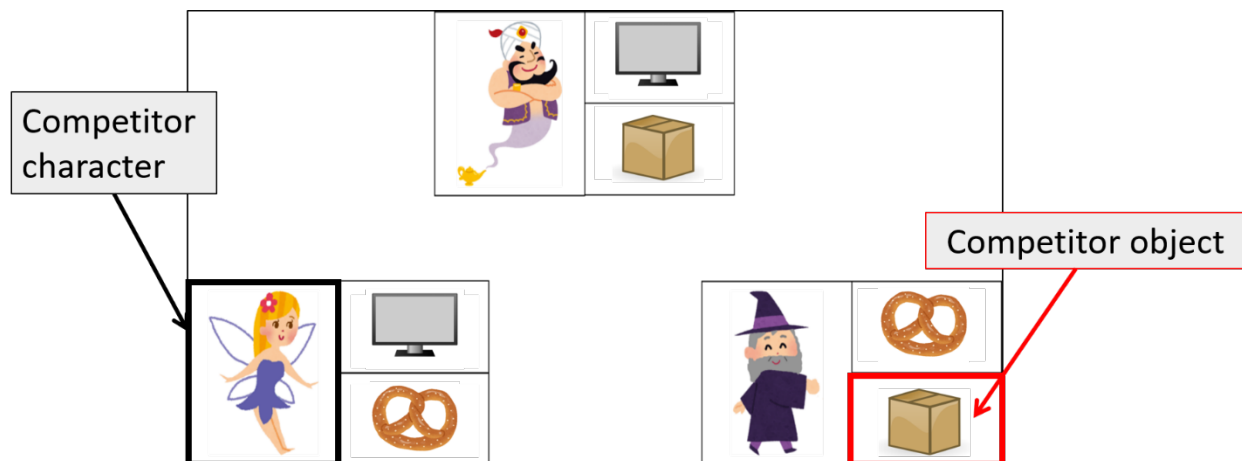


Figure 2: Visual display of trial for sentence (14) with competitors highlighted.

For the predicted eye-movement patterns (analyzed only for Condition 3), we refer to the target sentence for Condition 3, repeated in (14) above, and to Figure 2 which shows the visual display and highlights the two areas of interest. To evaluate the truth value of the target sentence, a participant must assess both the presupposition and the assertion related to *only* (Horn, 1969). For the presupposition to be true, the character must have gotten the target object; hence, participants are expected to launch more looks toward the target object (*the pretzel*) to evaluate the presupposition. For the assertion to be true, it must be the case that the target character is the only character in the visual scene who possesses the target object; i.e., no other character in the discourse context should have received a pretzel. Because the discourse context for Condition 3 falsifies the assertion, it is expected that an adult-like pattern of fixations will yield increased

looks to the competitor character who logically falsifies the sentence (e.g., *the Fairy*), or potentially to the other object in the visual scene that matches the target object (e.g., *Fairy's pretzel*).

A different pattern of eye fixations may be expected for children who commit a spreading error and interpret the focus operator as taking scope over the VP. A VP-scope analysis requires that no other objects were received by target character. This scope analysis makes clear predictions for eye-movement data, such that children are *not* expected to search the visual scene for other characters who received the target object, but instead to search the set of objects received by the target character to evaluate whether only one object was received. Thus, under a spreading error, it is predicted that children will show increased fixations to a competitor *object* (e.g., *the box*) rather than the competitor *character*. However, as in Zhou et al. (2012), it is also possible that eye movement data will reveal qualitatively similar processing patterns for adults and children while resulting in similar or different judgment patterns.

Returning to the manipulation of QAC, it is proposed that following results from Hackl et al. (2015), QAC-congruent and -incongruent questions will yield different truth value judgments and eye gaze patterns across the trials. Specifically, it is expected that children in the Experimental group will yield more adult-like eye movement patterns as compared to children in the Control condition. An asymmetry in processing patterns would provide further evidence that children's analysis of the focus operator is highly subject to processing demands via the pragmatic felicity of the context question. The following section reports results for all participants. To establish a baseline, the adult results are outlined first.

4. Results

4.1. Adult results. Overall, adults were highly accurate in their truth-value judgments, scoring 96.4% overall and 97.7% on the target subject-*only* sentences. As predicted, adults did not differ across groups in their accuracy, scoring nearly perfect on the subject-*only* sentences for both the control group (99.2%) and experimental group (96.2%). In sum, adults performed with near-ceiling accuracy, reflecting that they correctly computed the interpretation of subject-*only* across conditions. Next, we turn to adults' eye movement patterns for the critical Adult-*false* / Child-*false* Condition 3.

In the best-fitting model for the first time window (the segment containing "*only*"), a significant effect of Group ($t(141) = 2.862, p < 0.01$) as well as a significant interaction of Group \times AOI ($t(141) = -2.987, p < 0.01$) emerged. To interpret the interaction of Group \times AOI, follow-up analyses were carried out by creating separate models for the two levels of Group. In the model for the Control group, the fixed effect for AOI was not significant ($t(67) = 0.890, p = 0.377$), indicating that there was no statistical difference in mean fixations to the competitor character and the competitor object for this group. In the model for the Experimental group, the AOI term was significant ($t(75) = -3.190, p < 0.01$). This indicates that the Experimental group made significantly more fixations to the competitor character as compared to the competitor object.

In the second time window (containing the mentioned character, e.g., "*the Wizard*"), the same pattern of results was evidenced: in the overall best-fitting model, there was a significant effect of Group ($t(151) = 2.458, p < 0.05$) as well as a significant interaction of Group \times AOI ($t(151) = -3.448, p < 0.01$). In the follow-up analysis investigating the interaction term, the model for the Control group revealed that the fixed effect for AOI was not significant ($t(75) = 0.713, p = 0.48$), whereas in the model for the Experimental group, the AOI term was significant ($t(77) = -3.881, p < 0.001$). As in the first segment, in Segment 2, the Control group showed equal fixations to the competitor character as to the competitor object, while the Experimental group showed significantly more fixations to the competitor character than the competitor object.

In the third time window (containing "*got*"), the overall best-fitting model revealed a significant effect of Group ($t(137) = 2.963, p < 0.01$), AOI ($t(137) = 2.003, p < 0.05$), and an interaction of Group \times AOI ($t(137) = -3.548, p < 0.001$). In light of the interaction term, prior to evaluating the fixed effects two follow-up models were conducted for each Group. The Control group model revealed a significant effect of AOI ($t(71) = 2.128, p < 0.05$), indicating that the Control group made more fixations to the competitor object than the competitor character. In the follow-up model for the Experimental group, AOI was also significant ($t(67) = -2.854, p < 0.01$). In this group, there were significantly more fixations to the competi-

tor character as compared to the competitor object. Thus, the Control and Experimental groups showed significantly different patterns of fixations in the third time window, with the Control group looking more at the competitor object and the Experimental group looking more at the competitor character.

For the final time window in the sentence (containing the mentioned object, e.g., “a pretzel”) the overall best-fitting model revealed a significant effect of Group ($t(163) = 4.023, p < 0.001$), AOI ($t(163) = 3.233, p < 0.01$) and an interaction of Group \times AOI ($t(163) = -5.263, p < 0.001$). In light of the interaction term, prior to evaluating the fixed effects, two follow-up models were conducted for each level of Group. The Control group model revealed a significant fixed effect of AOI ($t(85) = 3.567, p < 0.001$), indicating that the Control group made more fixations to the competitor object than the competitor character. In the follow-up model for the Experimental group, the AOI term was also significant ($t(79) = -3.818, p < 0.001$). In this group, there were significantly more fixations to the competitor character as compared to the competitor object. As in the previous time window, the Control and Experimental groups showed significantly different patterns of fixations in the third time window, with the Control group fixating more at the competitor object and the Experimental group fixating more at the competitor character.

To summarize the results from all four time windows, eye-movement data revealed a robust QAC effect (see Figure 3). As predicted, the Experimental group made more fixations to the ‘correct’ competitor character as early as the first time window. On the other hand, the Control group looked at the two competitors equally for the first two time windows; in the third and fourth time windows, this group made more fixations to the ‘incorrect’ competitor object, which would falsify the sentence under a spreading interpretation. These results reveal an important role for discourse pragmatics in the processing of subject-*only* for adults. When the discourse context was pragmatically felicitous with the target sentence (i.e., the QAC-congruent Experimental group), adults’ processing was facilitated, as indexed by increased fixations to the correct falsifying character. This pattern is in line with our predictions for successful processing of subject-*only* in this Adult-*false*/Child-*false* condition. In contrast, when the discourse context was pragmatically incongruent with the target sentence (i.e., QAC-incongruent Control group), adults made more fixations to the incorrect falsifier, the competitor object. This pattern suggests that adults in the Control group process the subject-*only* target sentence exhibiting what would take place for ‘spreading error’ interpretation, as we predicted would be the case for children. Importantly however, adults in the Control group performed at ceiling in their end-of-sentence truth-value judgments, indicating that their momentary consideration of the incorrect falsifier did not lead them to ultimately misinterpret the subject-*only* sentence, as is expected for children. We return to this distinction in the Discussion.

4.2. Child results.

4.2.1 TVJ Accuracy. Behavioral accuracy on the truth value judgment task across all conditions (including control and fillers) for children was 78%. However, TVJ accuracy to the three subject-*only* conditions was lower, at 60%. To further examine children’s comprehension of the target conditions, accuracy was examined separately per group. Table 6 shows accuracy per condition separately per group.

Condition	Control group ($n=6$)	Experimental group ($n=7$)
1. Adult <i>true</i> / Child <i>false</i>	58.3	41.7
2. Adult <i>false</i> / Child <i>true</i>	37.5	45.8
3. Adult <i>false</i> / Child <i>false</i>	79.2	100

Table 6: TVJ accuracy per condition per group (%).

Examining each condition separately revealed variability across groups. As the focus of the current paper is the Adult-*false*/Child-*false* Condition 3, we focus here on accuracy in that condition. Given that the correct answer is ‘false’ to target sentences in this condition, the raw accuracy values in Table 6 do not provide a comprehensive picture of accuracy, given that the target sentence is considered ‘false’ under

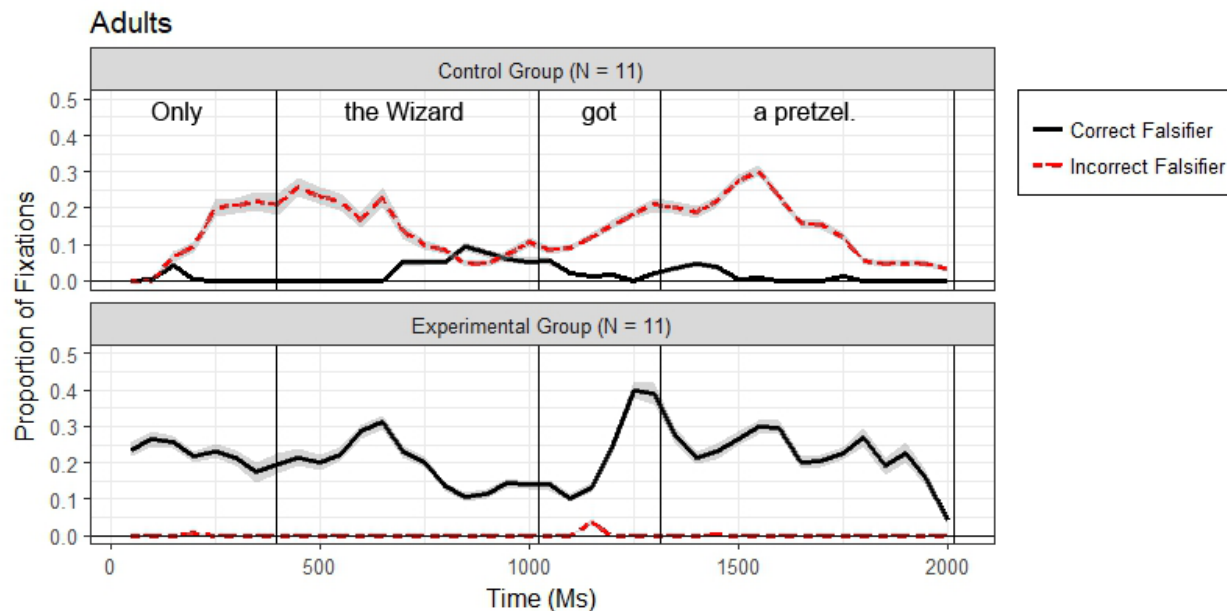


Figure 3: Proportions of eye fixations during sentence presentation to the correct (black) and incorrect (red) falsifiers for adults.

both a correct/adult interpretation, and a spreading/child interpretation. Thus, follow-up justification responses are important to help tease apart whether the child supplies an adult-like justification (in which the competitor character was selected as the falsifier) or whether a non-adult-like justification is provided (which would mention the competitor object as the falsifier).

Of the correct ‘false’ responses (97% of responses across children), 53% of the responses were followed-up with an adult-like justification in which the competitor character was mentioned. In other words, for 47% of trials in Condition 3, children provided a non-adult-like justification, mentioning the ‘incorrect’ competitor object which did not falsify the target sentence. For the Control group, non-adult-like justifications were given for 50% of trials in this condition; for the Experimental group, non-adult-like justifications were given for 44% of trials. Thus, about half of ‘false’ responses in each group were followed-up with non-adult-like justifications.

Thus far, this pattern is not in line with predictions of Hackl et al. (2015), in which pragmatic felicity is predicted to increase children’s accuracy to subject-only. One reason why the emerging results do not replicate Hackl et al.’s findings may be the small sample size; we leave this as an open question, addressing the need for further testing (which we are currently conducting). To examine online processing, we turn next to children’s eye-fixation patterns. In the interest of preserving as much data as possible, all analyses reported below include all trials across participants, independent of whether a child supplied an adult-like or non-adult-like interpretation. In addition, given the exploratory nature of this statistical analysis and the size of the dataset, we made the decision to run separate models for each Group (Control, Experimental) rather than including Group as a fixed effect within a larger model. This allowed us to more robustly analyze group differences given the relatively small group sizes.

4.2.2 Eye-movement data.

CONTROL GROUP. In the model for the first time window, the fixed effect of AOI was not significant ($t(33) = -0.603, p = 0.55$), indicating that there was no statistical difference in mean fixations to the competitor character and the competitor object. This was also the case for the second ($t(37) = -1.015, p = 0.32$), third ($t(37) = -0.942, p = 0.35$), and fourth time windows ($t(39) = 0.348, p = 0.73$).

EXPERIMENTAL GROUP. In the model for the first time window, the fixed effect of AOI ($t(47) = -1.402$, $p = 0.17$) was not significant, indicating that there were an equal number of fixations made to the competitor character and the competitor object in the first segment. The model for the second time window revealed a significant effect of AOI ($t(43) = -2.277$, $p < 0.05$), indicating that there were more fixations made to the competitor character than to the competitor object in this segment. For both time windows three and four, the effect of AOI was not significant (Segment 3: $t(35) = -1.264$, $p = 0.21$; Segment 4: $t(47) = 0.237$, $p = 0.81$).

To summarize results across groups, eye-movement data revealed group differences in fixations in the second time window (see Figure 4). In the second time window (during auditory presentation of the mentioned character, e.g., “the Wizard”), the Experimental group made more fixations to the ‘correct’ competitor character, indicating that pragmatic support in the preceding prompt question resulted in this group having a processing advantage. Note that, however, this significant difference did not extend for the subsequent two time windows. In contrast, the Control group looked at the two competitors equally across the presentation of the entire sentence. These results are discussed in what immediately follows.

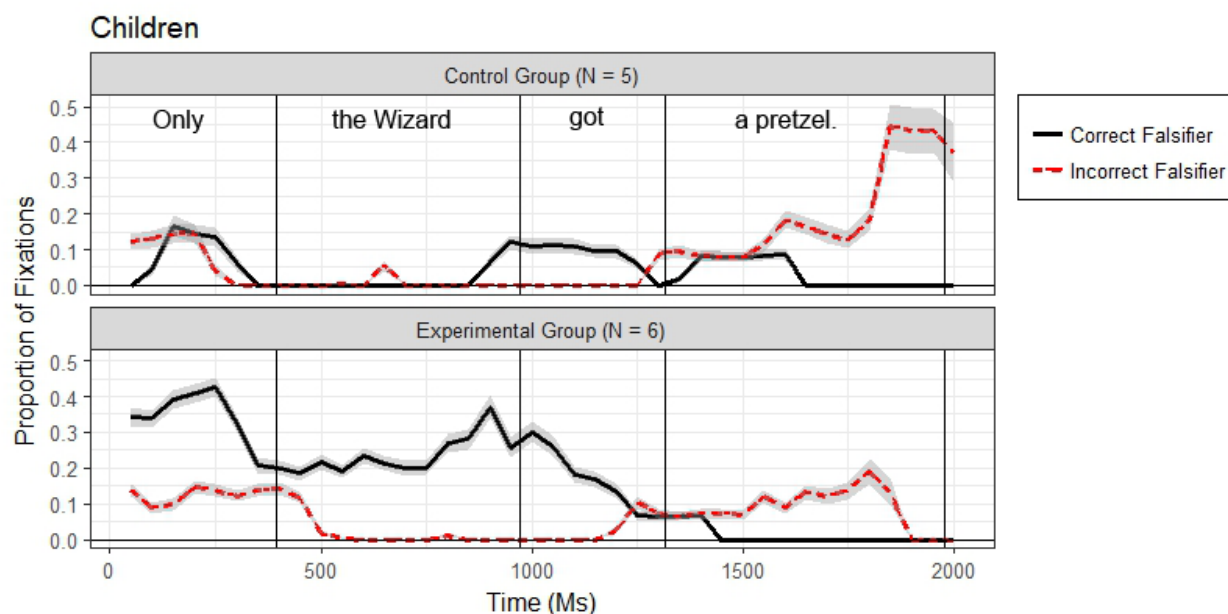


Figure 4: Proportions of eye fixations during sentence presentation to the correct (black) and incorrect (red) falsifiers for children.

5. Discussion

To summarize the results, we begin with the adult participants, who performed at ceiling in their meaning comprehension responses. Interestingly, adults' online processing was modulated by pragmatic felicity, such that the experimental group was found to have increased fixations to the logically falsifying competitor character, whereas the control group appeared to make increased fixations to the ‘incorrect’ competitor, the non-adult-like falsifying object. These patterns provide some of the first preliminary evidence that adults' online processing of subject-*only* is modulated by Question-Answer Congruency of the discourse context, in line with the adult data from Sugawara (2016). Sugawara found that adults had difficulty comprehending subject-*only* sentences in a condition which did not have pragmatically felicitous prompt questions. This suggests that it is difficult even for adults to ignore the ‘incorrect’ falsifier without sufficient pragmatic support. While we did not observe differences in adults' offline responses, their eye-movement data suggest that rich pragmatics in the discourse resulted in processing facilitation for the QAC-congruent Experimental group, as indexed by increased fixations to the expected falsifying character. This further suggests that if pragmatic felicity plays an important role in adults' online processing of subject-*only*, then it may be equally

(or even more so) important in children's processing.

Turning to the children's data, our sample thus far did not reveal group differences in children's end-of-sentence judgments. For the critical *Adult-false/Child-false* Condition 3, both groups provided non-adult-like end of sentence justifications on about half of the trials. Interestingly, this pattern appears to be driven by two different 'responder types'. That is, the majority of children responded accurately on either all 4/4 trials or 0/4 trials; this was true for 8 of the 11 children. While this preliminary pattern suggests that children are either categorically adult-like or non-adult-like in their offline judgments, what remains unclear is which factors affect whether or not a child responds in an adult-like way. To address this open question, a next step of this study will analyze data from a variety of individual difference measures, including tests of memory, executive control, and grammatical knowledge. This will allow us to determine whether (and how) certain cognitive or linguistic abilities may predict children's comprehension of *only*.

Children's eye-movement data revealed an interesting pattern of results which were similar to those of the adult groups. Processing in the Experimental group appeared to be somewhat more adult-like, with this group showing increased fixations to the logically falsifying character. This trend is taken to suggest that pragmatic felicity of the context question allows children to more strongly consider the adult-like competitor character during processing. In other words, as compared to children in the Control group, who showed no differences in fixations to the two potential competitors, the Experimental group appeared to have a processing advantage. This directly parallels results found for adults, with the experimental group showing increased fixations to the competitor character. While we were unable to include a direct age comparison of adult and child eye-movement data due to imbalance in sample sizes, this similarity suggests that pragmatic felicity is at play in both children and adults for the *Adult-false/Child-false* condition. Future analyses directly comparing adults with children who provided adult-like responses are called for to evaluate and compare children's processing in terms of how they reach the adult-like computation.

Together, adult and child data suggests that pragmatic richness in the discourse plays an important role during meaning comprehension of subject-*only*. This finding is in line with results from [Hackl et al. \(2015\)](#) and [Sugawara \(2016\)](#), who argued that Crain's puzzle is better explained by processing issues relating to pragmatic infelicity. Note that results from the child Control group, who were presented with pragmatically infelicitous sentences, do not thus far show clear disadvantages for interpretation or processing above and beyond the Experimental group, which is expected straightforwardly under an account which posits processing costs for pragmatic infelicity. However, what is suggested thus far is that initial findings may not be directly compatible with accounts attributing children's errors to a non-adult-like aspect in the grammar, such as the Sentence Analysis Hypothesis ([Notley et al., 2009](#); [Zhou & Crain, 2010](#)), which would require further clarification regarding why children are sensitive to the discourse. Given that both the Experimental and Control groups were presented with the same test sentences (which, crucially, did not have an overt element to block the c-command-driven 'spreading' of *only* from the pre-subject position to the VP, i.e., a Relativized Minimality effect), such an account would not predict differences to emerge between control and experimental groups.

6. Conclusion

In conclusion, emerging findings from the present study reveal an interesting role for pragmatic felicity in the processing of sentences containing subject-*only*. Using a sentence-story verification task in tandem with the visual world eye-tracking paradigm, the current study revealed that pragmatic richness in the discourse affected processing in both adults and children, particularly in conditions which were logically false. This difference was realized as a difference in fixations toward the character in the context who logically falsified the target sentence. Specifically, adults and children in the pragmatically congruent experimental condition made increased fixations toward the falsifying character, which further suggests that pragmatic felicity resulted in increasingly adult-like processing for children. Results thus far provide support for the Pragmatic Felicity Hypothesis, and imply, in line with this account, that previous studies examining Crain's puzzle may have underestimated children's ability to interpret subject-*only* due to the pragmatic infelicity of the prompt questions used in these experiments. Further research is called for to strengthen the current findings and suggestions we can draw from them.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. doi: arXiv:1406.5823.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Crain, S., Ni, W. J. & Conway, L. (1994). Learning, parsing, and modularity. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 443–467). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crain, S., Philip, W., Drozd, K., Roeper, T. & Matsuoka, K. (1992). Only in child language. Paper presented at the 17th Annual Boston University Conference on Language Development. Boston, MA.
- Crain, S., & Thornton, R. (1998). *Investigations in universal grammar*. Cambridge, MA: MIT Press.
- Endo, M. (2004). Developmental issues on interpretation of focus particles by Japanese children. In *BUCLD: Proceedings of the 28th Annual Boston University Conference on Language Development* (pp. 141–152). Somerville, MA: Cascadilla Press.
- Goro, T., Minai, U., & Crain, S. (2005). Two disjunctions for the price of only one. In *BUCLD: Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 228–239). Somerville, MA: Cascadilla Press.
- Goro, T., Minai, U., & Crain, S. (2006). Bringing out the logic in child language. In *Proceedings of the 35th Annual Meeting of the North East Linguistic Society* (pp. 245–256). Somerville, MA: Cascadilla Press.
- Hackl, M., Sugawara, A. & Wexler, K. (2015) Question-Answer (In)Congruence in the Acquisition of Only. In *BUCLD 39: Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 204–217). Somerville, MA: Cascadilla Press.
- Horn, L. (1969). A presuppositional theory of only and every. *CLS 5, Chicago Linguistics Society*.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. R package version, 2-0.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Müller, A., Schultz, P., & Höhle, B. (2011). How the Understanding of Focus Particles Develops: Evidence from Child German. In *Selected Proceedings of the 4th Conference on Generative Approaches to Language Acquisition North America (GALANA2010)* (pp. 163–171). Somerville, MA: Cascadilla Proceedings Project.
- Notley, A., Zhou, P., Crain, S. & Thornton, R. (2009). Children's Interpretation of Focus Expressions in English and Mandarin. *Language Acquisition*, 16, 240–282.
- Paterson, K. B., Liversedge, S. P., Rowland, C., & Filik, R. (2003). Children's comprehension of sentences with focus particles. *Cognition*, 89, 263–294.
- Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Niemeyer, Tübingen.
- Paul, P., Ziegler, J., & Snedeker, J. (2017). Incremental processing of only-sentences in adults and children. Poster presented at the 30th CUNY Conference on Human Sentence Processing, Boston, MA.
- Philip, W., & Lynch, E. (2000). Felicity, relevance, and acquisition of the grammar of every and only. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.
- Rizzi, L. (2001). Relativized minimality effects. In M. Baltin, & C. Collins (eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 89-110). Malden, MA: Blackwell Publishers Ltd.
- Rooth, M. (1985). *Association with Focus*. Doctoral dissertation, University of Massachusetts, Amherst.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1), 75–116.
- Sano, T. (2011). Abstract CP/IP configuration in child Japanese. In *BUCLD 36 Proceedings* (pp. 525–535). Somerville, MA: Cascadilla Press.
- Sugawara, A. (2016). *The role of question-answer congruence (QAC) in child language and adult sentence processing*. Doctoral dissertation, Massachusetts Institute of Technology.

- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.
- Zhou, P., & Crain, S. (2010). Focus identification in child Mandarin. *Journal of Child Language*, *37*, 965–1005.
- Zhou, P., Su, Y. E., Crain, S., Gao, L., & Zhan, L. (2012). Children's use of phonological information in ambiguity resolution: A view from Mandarin Chinese. *Journal of Child Language*, *39*, 687–730.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).