Erschienen in: Journal of Child Language ; 45 (2018), 2. - S. 415-441
https://dx.doi.org/10.1017/S0305000917000277

J. Child Lang. **45** (2018), 415–441.
doi:10.1017/S0305000917000277

On-line processing of English *which*-questions by children and adults: a visual world paradigm study*

CARLA CONTEMORI

Department of Languages and Linguistics, University of Texas, El Paso

MATTHEW CARLSON

Department of Spanish, Italian and Portuguese, Pennsylvania State University and Center for Language Sciences, Pennsylvania State University

AND

THEODOROS MARINIS

School of Psychology and Clinical Language Sciences, University of Reading, UK

(*Received* 1 *August* 2016 – *Revised* 17 *January* 2017 – *Accepted* 28 *June* 2017 – *First published online* 25 *July* 2017)

ABSTRACT

Previous research has shown that children demonstrate similar sentence processing reflexes to those observed in adults, but they have difficulties revising an erroneous initial interpretation when they process garden-path sentences, passives, and *wh*-questions. We used the visual-world paradigm to examine children's use of syntactic and non-syntactic information to resolve syntactic ambiguity by extending our understanding of number features as a cue for interpretation to *which*-subject and *which*-object questions. We compared children's and adults' eye-movements to understand how this information shapes children's commitment to and revision of possible interpretations of these questions. The results showed that English-speaking adults and children both exhibit an initial preference to interpret an object-*which* question as a subject question. While adults quickly override this preference, children take significantly longer, showing an overall processing difficulty for object questions. Crucially, their recovery from an initially erroneous interpretation is speeded when disambiguating number agreement features are present.

[*] This work was supported by the Fondazione Marica De Vincenzi under the 2011 Postdoctoral Grant to the first author, which is gratefully acknowledged. Address for correspondence: Carla Contemori, Department of Languages and Linguistics, University of Texas at El Paso, 500 W. University Ave, El Paso, TX 79968. e-mail: carla.contemori@gmail.com

INTRODUCTION

The domain of *wh*-questions has been widely investigated in first language acquisition, mainly using production and off-line comprehension measures (e.g. elicitation, picture matching tasks), showing an asymmetry in children's interpretation of *wh*-questions: although children produce *wh*-questions already at age 1;7–3;0 (Guasti, 1996, a.o.), *who*-object questions pose a greater challenge than *who*-subject questions across languages, both in comprehension and production. Cross-linguistic research has also shown that among object extracted *wh*-questions, *which*-object are the hardest types of *wh*-questions to comprehend for children (for French: Jakubowicz & Gutierrez, 2007; for Greek: Stavrakaki, 2006; for Italian: De Vincenzi, Arduino, Ciccarelli & Job, 1999; for Hebrew: Friedmann, Belletti & Rizzi, 2009, a.o.), and English-speaking children do not reach full mastery of *which*-object questions until the age of 7 (for comprehension: Avrutin, 2000, for children between 3;5 and 5;2; Goodluck, 2005, for children aged 4–5; Deevy & Leonard, 2004, for children aged 2;9–6;10; Yoshinaga, 1996; for production: Stromswold, 1995). For instance, by using a picture pointing task, Avrutin (2000) found that English-speaking children aged 3;5 to 5;2 (mean age = 4;3) comprehend *which*-object questions less well than *which*-subject questions (48% correct vs. 86% correct responses), whereas such an asymmetry was not attested for *who*-questions (80% correct responses in both cases). Furthermore, Goodluck (2005) showed that the difficulty observed with *which*-object questions in a picture selection task was attenuated when a *wh*-phrase indicating a particular animal (e.g. 'which cow') was replaced by a more generic noun (e.g. 'which animal').

Various theoretical accounts have been proposed to explain the delay in the comprehension of object questions. In the present paper we present probabilistic models of sentence processing and use them to account for the subject/object asymmetry in *which*-questions comprehension (e.g. MacWhinney & Bates, 1989; Roland, Dick & Elman, 2007; Levy, 2008;). Under the probabilistic models of sentence processing, it has been suggested that the processing difficulty observed with object questions may result from an interruption of the comprehender's predictions (e.g. Roland *et al.*, 2007; Levy, 2008). For example, by looking at the comprehension of *wh*-questions in adults, expectation-based accounts postulate a processing cost at the point at which a rare syntactic structure is first encountered. In the case of an object dependency, the comprehender expects a more frequent subject extracted structure (see Diessel, 2004, for an analysis of child-directed speech), and when this expectation is not met, the result is a processing difficulty. For instance, in the sentence "Which dog did the cat bite?" the expectation-based accounts predict that, upon encountering the subject *the cat*, the parser would experience a slow-down. In the domain of child language processing, MacWhinney and Bates' (1989) Competition Model has suggested that children interpret sentences by using cues of varying strength, and the cues' reliability varies across languages. For instance, word order is a reliable cue to identify agents in English who are typically animate entities occurring preverbally in a sentence. As English-speaking children build their parsing preferences over time, the parser registers word order regularities and uses this cue reliably to guide interpretation. When the child parser's expectations about word order and animacy of the preverbal noun are not met, such as in the case of an object question, children may experience processing difficulties. For example, studies on the acquisition of object dependencies have observed that children have comprehension difficulties with some types of object dependencies (i.e. object relative clauses; Kidd, Brandt, Lieven & Tomasello 2007; Diessel, 2009), and they have linked this difficulty to cue reliability and frequency of occurrence of the structures in speech. For instance, Kidd *et al*. (2007) compared children's performance in elicited repetition with spoken corpus data, suggesting that children are sensitive to constraints on distributional contingencies present in the input. As a result, object relatives that are more frequently occurring in speech (e.g. where the object NP is inanimate and the subject NP is animate) are not as hard for children aged 3;1 to 4;9 to produce.

Although many studies have investigated the acquisition of filler–gap dependencies, such as *wh*-questions, very few have analyzed the real-time processing of these grammatical structures in children. Therefore, it is hard to tease apart the role that expectations based on word order and animacy may have in the delayed mastery of object questions. The first goal of the present study is to provide evidence on children's real-time processing of *wh*-questions. By exploring the contribution of incrementality during processing, we aim at understanding children's interpretation biases to contribute to current theories about language processing in child language acquisition.

The second goal of the study is to investigate adults' and children's revision abilities at different stages during the processing of similar sentence structures. To this aim, we test *which*-questions by manipulating Number features, as shown in (1)–(4), in order to provide early and late cues for interpretation.

(1) Which cows are pushing the goat?
(2) Which cow is pushing the goat?
(3) Which cow are the goats pushing?
(4) Which cow is the goat pushing?

Children are not expected to show problems in the interpretation of *which*-subject questions, such as (1)–(2), that are known to be relatively easy for children to comprehend. However, we expect them to present difficulties with *which*-object questions, such as (3)–(4), possibly at an increased level when the number features of the first NP and auxiliary are similar, because the disambiguation towards an object interpretation occurs later (i.e. *are* in (3); *the goat* in (4)). By manipulating number features in *which*-object questions, our study tests the processing of filler–gap dependencies where disambiguation is provided earlier (3) or later (4) in the structure. This will test more closely the commitment and revision abilities of children in comparison to adults.

Filler–gap dependencies and revision in child language acquisition

Subject (S-WH) and Object (O-WH) *wh*-questions are an excellent testing ground for the analysis of the incrementality and revision in processing routines in children compared to adults. In adult sentence processing, a dislocated constituent (or 'filler') is thought to trigger the anticipation of a lexical head to license it, or of a corresponding syntactic gap (Filled Gap effect; Stowe 1986; Frazier & Clifton, 1989; Frazier & Flores d'Arcais, 1989). For instance, Stowe (1986) hypothesized that adults anticipate a gap in the processing of sentences like (5) and (6), by observing slower reading times for the pronoun *us* following the lexical verb *bring* in a *wh*-fronting condition such as (5), relative to a control condition that did not involve *wh*-fronting, such as (6).

(5) My brother wanted to know who Ruth will bring us home to ____ at Christmas.

(6) My brother wanted to know if Ruth will bring us home to Mom at Christmas.

A few studies on the processing of filler–gap dependencies (*wh*-questions; relative clauses) in adults have shown a subject gap bias in processing filler–gap dependencies across languages (e.g. Schlesewsky, Fanselow, Kliegl & Krems, 2000, for German; Frazier & Flores d'Arcais, 1989, for Dutch; Levy, 2008; Staub, 2010, for English), as predicted under the probabilistic accounts of sentence processing (see also the NVN heuristic; Bever, 1970). For example, Staub (2010) tested the processing of object relative clauses in English-speaking adults using eye-tracking during reading and showed that at least part of the processing difficulty associated with object relatives occurs because they are less frequent than subject relative clauses. This effect of frequency was observed in an increased processing cost associated when participants encountered the subject NP within the relative clause, suggesting that because the parser was expecting a subject gap, the violation of this expectation resulted in a general processing difficulty with the structure. In the present study, we aim at analyzing the early stages of processing in *which*-object questions, to observe the early commitments in child language processing. Based on the findings from the adult literature, the child parser should also be sensitive to the frequency of subject questions, and expect a subject gap when an object question is processed (e.g. Kidd *et al*., 2007).

The second aim of the current study is to investigate the process of revision in the on-line comprehension of *which*-questions. The adult parser allows the comprehender to properly retract early incremental commitments to a subject interpretation. However, children may not always be able to recover from the interpretation bias because they do not have fully developed revision abilities. As previously observed in studies investigating the on-line processing of garden path sentences (Trueswell, Sekerina, Hill & Logrip, 1999), passives (Huang, Zheng, Meng & Snedeker, 2013), and *wh*-questions (Omaki, Davidson-White, Goro, Lidz & Phillips, 2014), the children's parser is not always as flexible as the adult parser. Children's difficulty with revising seems to be a general developmental phenomenon across languages (e.g. Novick, Trueswell & Thompson-Schill, 2005), and maturational differences have been used to explain why the first interpretation that children arrive at tends to be the only interpretation they can entertain.

In particular, the body of research looking at the processing of complex syntactic structures has long suggested that children's difficulty with revising is the result of the development of general executive function (EF) processes, specifically the ability to select competing representations (e.g. Trueswell & Gleitman, 2004; Novick *et al*., 2005; Novick, Kan, Trueswell & Thompson-Schill 2010; Woodard, Pozzan & Trueswell, 2016). Furthermore, more recent research has linked the ability to detect statistical regularities in the input (i.e. statistical learning) with the successful comprehension of complex syntactic structures, such as passives and object relative clauses (Kidd & Arciuli, 2016).

The study by Omaki *et al*. (2014) is an interesting example of the revision difficulty with filler–gap dependencies observed in children. Omaki and colleagues tested the off-line comprehension of *wh*-questions like (7) and (8a) in English- and Japanese-speaking children.

(7) Where did Lizzie tell someone that she was gonna catch butterflies?

(8) a. Doko-de Yukiko-chan-wa choucho-o tsukamaeru-to itteta-no?
where-at Yukiko-DIM-TOP pro butterfly-ACC catch- COMP was telling-Q
'Where was Yukiko telling someone that she will catch butterflies?'

In (7) and (8a) there are two possible gap positions because the locative *wh*-phrase can be attached either to the main clause VP ('Yukiko telling someone') or the embedded clause VP ('she will catch butterflies'). In Japanese, the embedded clause is the first to complete and both adults and five-year-old children associate the *wh*-phrase with the gap in the embedded clause (i.e. the location where Yukiko caught butterflies). Omaki *et al*. (2014) included a PP in the structure to specify the location of the embedded clause event (e.g. *kouen-de* 'park-at'), as illustrated in (8b), and showed that adults preferred to adopt the main clause interpretation (i.e. the location where Yukiko told someone), because the PP (*kouen-de* 'park-at') had already filled the syntactic position of the locative. In contrast, five-year-olds still provided the embedded clause interpretation, as in (7), showing that they were not able to override the preference of interpreting the fronted *wh*-phrase as a filler for the embedded clause VP.

(8) b. Doko-de Yukiko-chan-wa [kouen-de choucho-o tsukameru to] itteta-no?
where-at Yukiko-DIM-TOP pro park-at butterfly-ACC catch COMP was telling-Q
'Where was Yukiko telling someone that she would catch a butterfly at the park?'

The study by Omaki *et al*. (2014) illustrates evidence that supports the incrementality of filler–gap dependency processing in children and a possible difficulty with its revision. However, the study used an off-line comprehension task and did not provide evidence for the timecourse of *wh*-question comprehension in children. Furthermore, as Omaki and colleagues pointed out, maybe children were not successful in making use of the filled-gap cue in the unambiguous condition because the fronted *wh*-phrase in the stimuli was an adjunct. According to Omaki *et al*., the flexibility in VP-adjunction (i.e. there is no restriction on the number of adjuncts that can be attached to a VP) might have influenced the effectiveness of the filled-gap manipulation.

In the present study, we employ the 'looking while listening' technique in an on-line eye-tracking experiment to investigate the early commitments and the reanalysis process of *wh*-questions in adults compared to children. To our knowledge, this is the first study that looks at this type of dependency using eye-tracking in a visual word paradigm. This technique provides a continuous record of the listeners' eye-movements and expectations as the utterance unfolds, yielding a detailed picture of child and adult processing over time, as well as an indication of the timecourse of initial expectations versus readjustment of expectations and recovery from misanalyses

(Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy 1995, a.o.). In particular, we will look at how and when children assign the correct interpretation to the filler in real time compared to adults, as well as how and when disambiguating information is being used for reanalysis, in order to give a correct interpretation to the sentence.

The role of number features in the processing of which-*questions*

In our investigation of the active filler–gap strategy in adults and children, we focus on the role of Number agreement in *which*-questions, by investigating the comprehension of *which*-subject questions with different number on the *which*-phrase (plural) and the auxiliary verb (singular), as shown in (9) (S-WH PS); with same number on the *which*-phrase (singular) and the auxiliary verb (singular), as shown in (10) (S-WH SS); *which*-object questions with different number on the *which*-phrase (singular) and the auxiliary verb (plural), as shown in (11) (O-WH SP); and similar number on the *which*-phrase (singular) and the auxiliary verb (singular), as shown in (12) (O-WH SS).

(9) S-WH-PS: Which cows _ are pushing the goat?
(10) S-WH-SS: Which cow _ is pushing the goat?
(11) O-WH-SP: Which cow are the goats pushing _?
(12) O-WH-SS: Which cow is the goat pushing _?

O-WH questions, such as (11) and (12), represent interesting cases of filler–gap dependencies because the Number feature on the auxiliary potentially disambiguates for an object interpretation and can address early and late effects of incremental processing. When presented with a sentence, such as (11), the number difference between the first NP (*cow*) and the auxiliary verb (*are*) is a strong syntactic cue that the filler cannot be posited yet, and thus that this is not a subject question. However, in the case of Number feature match between the first NP and the auxiliary, such as in (12), the interpretation of the sentence is consistent with that of both a subject and an object question until the parser encounters the subject NP (*the goat*), at which point it becomes clear that this is an object question. Therefore, in sentences such as (12), parsing may be harder as the duration of the ambiguity increases (e.g. for adults: Christianson, Hollingworth, Halliwell & Ferreira, 2001). If the strategy adopted by children and adults when interpreting the O-WH-SS is to commit to a subject interpretation at the upcoming verb (*is*), the listeners may take longer to make a revision when they encounter the subject (*the goat*) in an O-WH-SS compared to an O-WH-SP, because the number feature of the auxiliary in the O-WH-SP provides an early disambiguation cue, which is not available in the O-WH-SS. Concerning this effect, we may observe on-line facilitation for O-WH-SP compared to O-WH-SS in children. Children may take longer to revise O-WH-SS sentences, while they may resolve the ambiguity faster upon reaching the NP following the auxiliary in O-WH-SP sentences. In such cases, we could conclude that the children's processing revision of object questions is less flexible, and that their parser is more likely to fail to retract incremental commitments than adults (e.g. Trueswell *et al.*, 1999, Trueswell & Gleitman, 2004). Furthermore, if children are impacted to a greater degree by the later onset of disambiguating information in O-WH-SS sentences, they should be less accurate in correctly interpreting (12) compared to (11), whereas adults, who can recover more quickly, should be equally accurate.

A second possibility is that both adults and children will benefit equally from the presence of different number in the *which*-phrase and the auxiliary in O-WH-SP as compared to O-WH-SS sentences, and no difference in revision is observed across conditions in either group. If no clear effect of number is found in the on-line data in children compared to adults, it is still unclear if children will show different accuracy in the off-line comprehension of the two types of object questions.

Feature dissimilarity in filler–gap dependencies has been analyzed in adult parsing within cue-based memory retrieval models (e.g. Gordon, Hendrick & Johnson, 2004; Van Dyke & McElree, 2006), which proposed that sentences are represented as a group of small constituents that are retrieved through a cue-based search. During processing, the cue-based search allows the integration of words into the existing interpretation, but decay and retrieval interference can limit the parser's performance. For example, in the processing of a *wh*-question (e.g. *Which cow did the goat chase?*), these models hypothesize that upon encountering the verb (*chase*), a cue-based search should be activated, looking for constituents that have syntactic or semantic features that match the verb's selection properties (e.g. [+NP] or [+animate]). When the features of a word are closely related to the retrieval cues, then the word can be retrieved for additional processing. However, if the filler (*cow*) matches the cue-based search, and the intervening NP (*the goat*) also matches the search criteria, the intervening NP can interfere in the retrieval process of the filler at the gap position. In these models, the concept of retrieval interference is crucial in the interpretation of filler–gap dependencies, and explains the perceived complexity of these types of structures. Additionally, cue-based memory retrieval models make specific predictions about the processing of a filler–gap dependency, suggesting that the interference effects should be localized at the verb region where the filler-retrieval occurs. In the present study, if children are more affected by similarity-based interference in the processing of *which*-object questions with different number features, cue-based memory retrieval models would predict that interference effects should be localized to the verb region where the filler-retrieval occurs (e.g. the goat pushing _). Additionally, they would predict facilitation in the off-line comprehension of (11) in comparison to (12) in children.

To sum up, the questions we seek to answer about children's processing of *which*-questions are the following: (1) Is the early stage of processing for *which*-questions comparable in children and adults? (2) Do children's eye-movements show evidence of recovery from an erroneous subject interpretation of *which*-object questions (types (11) and (12)), or do they remain committed to the subject interpretation in the face of disambiguating information? Based on the age range selected for study (5;0–7;10), we expect to find evidence of recovery, but we do not yet know how this recovery plays out in the eye-movement record, and documenting the timecourse of children's processing of *which*-questions is thus a primary goal of this study.

In our study we focus on children aged 5;0 to 7;10. According to previous studies, children younger than 5;0 exhibit chance performance on the comprehension of *which*-object questions (e.g. Avrutin 2000). Furthermore, it is unclear at what age English-speaking children achieve adult-like performance on *which*-object questions (e.g. Deevy & Leonard, 2004, had only one group of children with a wide age-range of 2;9–6;10). We selected children older than 5;0 because at this age they should comprehend *which*-object questions more accurately than chance, and we recruited children up to the age of 7;10 to document the acquisition of these structures until a later stage of language development. Furthermore, to explore the effect of age on the processing of *which*-questions, we included age as a predictor in the analysis.

METHOD

Participants

Thirty-one English-speaking children and twenty-one adults participated in a visual-word-paradigm task. The adults were aged nineteen to forty-three (M = 31; SD = 12) students and graduate students at the University of Reading who were granted credit for their participation. The children were aged 5;0–7;10 (mean age: 6;04; SD: 0·10) and were randomly selected from the child development database at the University of Reading. All children and adults were monolingual speakers of English and did not have a history of language delay or impairment. The project received ethical approval from the University of Reading Research Ethics Committee. Data from four additional participants (1 adult and 3 children) showed severe track loss in the eye-tracking records (more than 40% of their data), and were discarded.

Materials

Participants performed a picture-selection task while their eye-gaze was tracked. Two experimental factors on *which*-questions were manipulated, Question type and Number, and each factor had two levels: Number (same number between the first NPs in the sentence and the auxiliary vs. different number between the first NP in the sentence and the auxiliary) and Question type (*which*-subject vs. *which*-object). A total of four conditions were tested, as shown in (9)–(12) above. Two of the four conditions, i.e. (9) and (11), display a similarity in number features between the first NP and the auxiliary, and two conditions, i.e. (10) and (12), display a dissimilarity in number features. Ten sentences were included per condition, which gave rise to forty experimental sentences.

All sentences were semantically reversible, and all nouns in subject and object NPs were animal names and matched in size in order to prevent any size-bias interpretation. Ten transitive verbs were selected (*spray, kick, splash, tickle, carry, kick, chase, kiss, stroke, push*) and each of them was used for two sets of questions. The nouns and verbs used in the sentences had an age of acquisition of five years or below based on the MRC Psycholinguistic Database (Wilson, 1988). Frequency of the animal nouns was measured with the English CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995). No statistical difference emerged between the frequencies of animal nouns presented in the sentences. Fillers were not included in the children's experiment due to time constraints. The adults' experiment also did not include fillers, to keep the sets of materials identical across age groups.

The experimental sentences were digitally recorded by two native speakers of English (a male and a female)[1] in a noise-proof sound booth. Sentences were randomly assigned to the two talkers. Each sentence was matched with two pictures, one of which appeared on the left side of the screen and one of which appeared on the right side. In the S-WH-SS and O-WH-SS conditions, one of the pictures showed a figure carrying out an action on another, while the other picture showed the same figures with the roles reversed. For the conditions with dissimilarity of number features (S-WH-PS and O-WH-SP), one character was carrying out an action on two other figures and the other picture showed the same figures with the roles reversed. To our knowledge, this is the first visual word study in which the pictures presented to the participants do not depict single referents (e.g. Trueswell *et al.*, 1999), but propositional meanings. An anonymous reviewer pointed out that a four-picture design with pictures of possible referents would have been more appropriate. We decided to

[1] Two different voices were used to record the sentences (half and half) in order to create variability in the voices and avoid boredom.

adopt two pictures because a four-picture design would have required an introductory story and two possible referents for the first NP (e.g. a white cow and a black cow), which may have added a level of complexity resulting in less clear looks to the target picture. However, depicting more complex propositional meanings, as we do here, means that we are not able to say what aspects of the pictures participants are attending to and, moreover, the pictures imply only two possible meanings intended by the talker, unlike in natural speech comprehension, where listeners begin with a larger and more diverse range of possible meanings. Figure 1 shows an example of each of the four conditions with the two corresponding pictures.

The sentences were pseudo-randomized so that items belonging to the same condition or containing the same nouns/verbs were counterbalanced in two blocks. Each participant encountered both blocks in two sessions in different order. Picture position on the screen was also counterbalanced, so that for half of the questions the correct picture appeared on the left side of the screen, and for the other half of the questions the correct picture appeared on the right side.

Procedure

To measure participants' eye-gaze, a Tobii T60 eye-tracking system (Tobii Technology AB, Sweden) was used, which tracks eye position every 16·7 ms with a resolution of 60 Hz. The remote Tobii eye-tracker is integrated in a 17-inch TFT monitor, has no visible or moving tracking devices, and allows a freedom of head movement of 44 × 22 × 30 cm. The eye-movement data reported are an average of both eyes. Stimulus presentation and eye-gaze data collection was conducted using E-prime (Schneider, Eschman & Zuccolotto, 2002). The testing started with a 5-point calibration procedure using Tobii Studio. The experimenter (first author) judged the quality of the calibration by examining the calibration plot for the five points. If the calibration accuracy did not meet the default criteria of the Studio software, the calibration was repeated.

The participants were tested individually in a quiet room of the laboratory. They sat on a chair at about 60 cm from the screen. A fixation cross was displayed on the screen prior to the start of each trial for 100 ms. After the fixation, participants saw two pictures on a screen for 2000 ms, then they listened to a question, while the two pictures remained on the screen. The auditory sentences were presented through headphones. After listening to the sentence, the participants had to choose the picture that correctly answered the question by pressing a button on a mouse. Participants were instructed at the beginning of the session that the right button corresponded to the right picture and the left button corresponded to the left picture. No time limit was provided for the comprehension question.

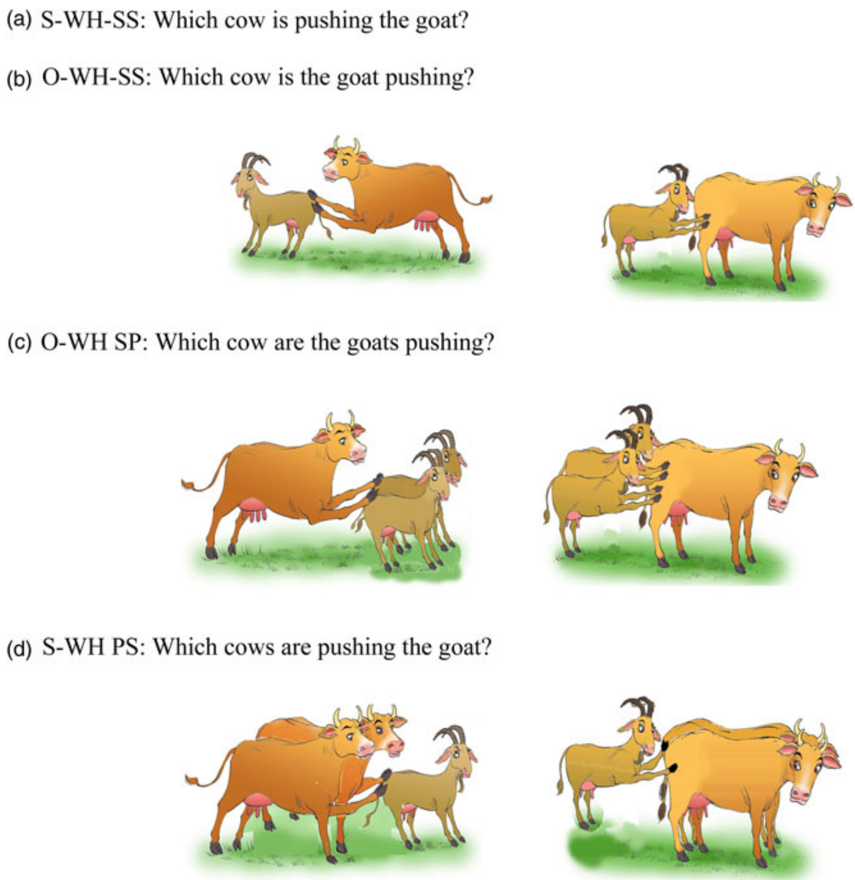

Fig. 1. Sample material.

At the beginning of the experiment, there was a practice block of five trials. This was to familiarize participants with the task. The practice items were also *wh*-questions similar to those used in the experiment. No feedback was provided by the experimenter on the participant's accuracy, but the experimenter explained the task again in cases where the participant had not understood.

The experiment was administered to both children and adults in two sessions, lasting about 10 minutes each. In the same session, children and adults completed another eye-tracking task and children completed a battery of off-line tests.

Data analysis

Comprehension data were recorded with a mouse and automatically coded as correct or incorrect, and analyzed using mixed-effects logistic regression, implemented in the lme4 package in R (Bates, Maechler, Bolker & Walker, 2015, version 1.1-10). The binary response served as the dependent variable, and significance of effects was tested here and for the eye-movement data via Likelihood Ratio Tests comparing the full model with models from which individual effects were removed.

Eye-movements were recorded by the E-prime software employing the standard settings. The output data we used for further analysis consisted of one text file per participant providing information about the specific timecourse of the experiment (e.g. onset of each trial) and the accurate position of the eye-gaze (as X–Y coordinates) at each time-point, as well as the number and duration of fixations according to E-prime default settings.

To analyze the looking behavior in relation to the verbal and visual stimuli presented, we defined two spatial areas of interest (AOI). Each AOI was 400 × 286 pixels in size, corresponding to the size of each of the pictures presented on the monitor. Eye-movements were timelocked to the onset of the auxiliary verb (*is*/*are*). The eye-movement data were analyzed starting from 200 ms after the onset of the auxiliary verb to account for the time it takes to program a saccadic eye-movement (Matin, Shao & Boff, 1993), and ending 1800 ms after the onset of the auxiliary. Trials with combined total looking times to the competitor and target of less than 30% of the trial duration (i.e. the 200–1800 ms following the auxiliary) were discarded, amounting to 3% of the data.

The period from 200 to 1800 ms after onset of the auxiliary was divided into 100-ms windows, and for each time-window, we calculated the proportion of looks to the competitor and target pictures, aggregated by condition for each participant. The proportion of looks to target, and the proportion of looks to the competitor, were analyzed separately, since they measure slightly different, though closely related, constructs, namely the ability to zero in on the target vs. the propensity to be distracted by the competitor, respectively. Since there were only two pictures on the screen, these proportions are of course nearly mirror images, but since fixations could also occur to locations outside of either AOI, they are not identical (40% of windows overall had proportions of looks outside the two AOIs in excess of 10%). The empirical logit transformation was applied to the resulting proportions, and analyzed using mixed-effects regression, weighted as described in Barr (2008), and implemented using lme4. Separate analyses were performed for looks to target and looks to competitor.

RESULTS

Accuracy results

Table 1 shows the proportion of correct responses given by the two groups in the four conditions. The adults were essentially at ceiling (see Table 1). Therefore, only the child data were analyzed statistically (although, interestingly, the only condition where the 95% Confidence Interval (CI) for the adults did not include 100% was the O-WH-SS, which we predicted to be the most difficult condition). The fixed effects were Question type (*which*-subject vs. *which*-object) and Number (same number between first NP and auxiliary in the sentence vs. different number between first NP and auxiliary in the sentence), and the child's age in months (*z*-scored). Question type and Number were sum-coded as $-0{\cdot}5$ (*which*-subject, and same number) and $+0{\cdot}5$ (*which*-object, different number). All interactions were allowed. The maximal, uncorrelated random effects structure by Participants and by Items was used. Items were considered to be the quadruplet of sentences involving the same animals and actions, since the differences in number, and in the thematic roles of each animal, are almost completely captured by the fixed effects.

The results confirmed the pattern of results shown in Table 1. There were significant main effects for Question type ($\beta = -2{\cdot}18$, $SE = 0{\cdot}45$, $\chi^2(1) = 17{\cdot}70$, $p < {\cdot}0001$) and Number ($\beta = 0{\cdot}94$, $SE = 0{\cdot}35$, $\chi^2(1) = 6{\cdot}07$, $p = {\cdot}01$), but the interaction between Question type and Number ($\beta = -2{\cdot}44$, $SE = 0{\cdot}74$, $\chi^2(1) = 7{\cdot}99$, $p = {\cdot}005$) showed that this was almost entirely driven by the lower accuracy rate for *which*-object questions with the same number (O-WH-SS). We confirmed the structure of the interaction by fitting models to only the data with *which*-object sentences, and to only the data in which the number of the subject and object differed. These, respectively, showed significantly poorer accuracy in comprehension of *which*-object sentences with the same number compared to different number ($\beta = 2{\cdot}15$, $SE = 0{\cdot}33$, $\chi^2(1) = 19{\cdot}70$, $p < {\cdot}0001$), and significantly poorer comprehension of *which*-object compared to *which*-subject sentences, even when the number was different ($\beta = -0{\cdot}98$, $SE = 0{\cdot}47$, $\chi^2(1) = 4{\cdot}03$, $p = {\cdot}04$).

In terms of the children's age, only the three-way interaction of Question type, Number, and Age approached significance ($\beta = 0{\cdot}92$, $SE = 0{\cdot}55$, $\chi^2(1) = 2{\cdot}93$, $p = {\cdot}09$; for the main effect of Age and its two-way interactions all ps $> {\cdot}3$), suggesting that, if anything, the pattern of results described above may be somewhat stronger for younger children.

In the next two sections we present the eye-movement results for children and adults.

Eye-tracking results

Figures 2 and 3 show the proportions of looks to the Target and Competitor pictures, respectively, for each condition, for children and adults. The data

TABLE 1. *Proportion of correct responses to the off-line comprehension questions, with 95% confidence intervals computed via non-parametric bootstrap (Agresti, 2002)*

Conditions	Adults	Children
S-WHSS	.99 [.98, 1]	.96 [.94, .98]
S-WHPS	.99 [.98, 1]	.95 [.93, .98]
O-WHSS	.96 [.90, .99]	.63 [.52,.74]
O-WHSP	.98 [.97, 1]	.89 [.83, .94]

were analyzed using growth curve analysis, implemented as a weighted mixed-effects regression with the empirical logit of the proportion of looks within each time-window as the dependent variable, and weights computed as in Barr (2008), as described above. By modeling the proportions of looks to target (or competitor) as a (non-linear) function of time, growth curve analysis (see also Mirman, 2014) allows a finer-grained view of the timecourse of performance than analyzing proportions of looks within a predefined time interval.

We fit two models, one to the proportions of looks to the Target, and one to the looks to the Competitor. These sets of proportions are largely complementary, save for fixations that fell outside the AOIs for both Target and Competitor. The results of these models were nearly identical, and we therefore present the results together, pointing out the differences as relevant. All valid trials (where no track loss occurred and total looking proportions to the Target and Competitor exceeded .30) were included in this analysis, regardless of the answer to the comprehension question, because comprehension accuracy and the eye-tracking record are simply different ways of measuring the same underlying processes. Trials with incorrect responses thus constitute valid data for testing those hypotheses.

Based on visual inspection of the data, Time was coded using a restricted cubic spline (Harrell, 2001) with four knots. This is the simplest spline that would adequately capture the shape of the Time function in the most complex condition being modeled (adults processing object-WH items; see Figures 2 and 3). The three components of the spline were decorrelated using principal components analysis. The fixed effects were Question type and Number, sum-coded as above, and Group, sum coded with adults coded as $-0{\cdot}5$ and children $+0{\cdot}5$. All three of these factors (Question type, Number, and Group) were then mean-centered, due to imbalance in the data. All interactions were allowed. The interactions involving the Time components capture the influence of Question type, Number, and Group on the shape of the Time function.

As above, significance was assessed using Likelihood Ratio Tests of the full model with models with the effects of interest removed. Since Time

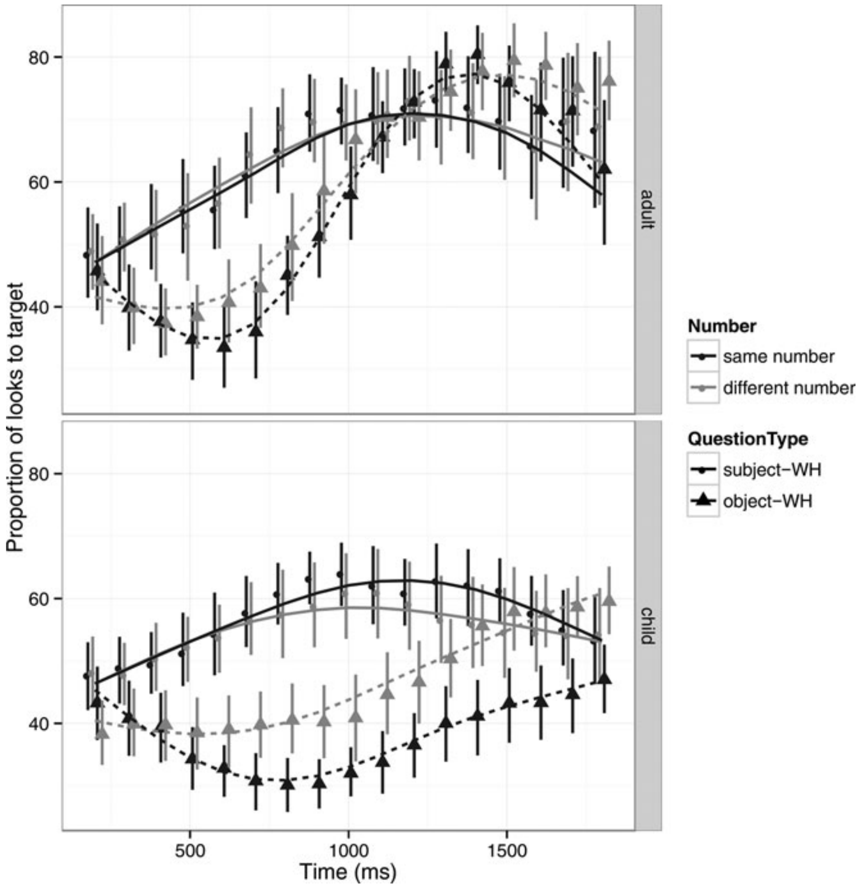

Fig. 2. Proportion of looks to the Target, by Group, Question type, and Number. Points and range bars show empirical means and 95% confidence intervals. Lines show model estimates.

was coded as a spline with four knots, the shape of the overall growth curve, and differences in the shape of that curve across experimental conditions (i.e. each interaction with Time), is described via three different coefficients, one for each component of the spline. For example, the two-way interaction of Time and Question type uses three coefficients (the interaction of each separate component of the spline with the Question type factor) to test how the shape of the growth curve varies between the two Question type conditions. Therefore, to test the Time by Question type interaction, all three of these terms were removed (i.e. three degrees of freedom), and the resulting model compared to the full model. Random by-Participant

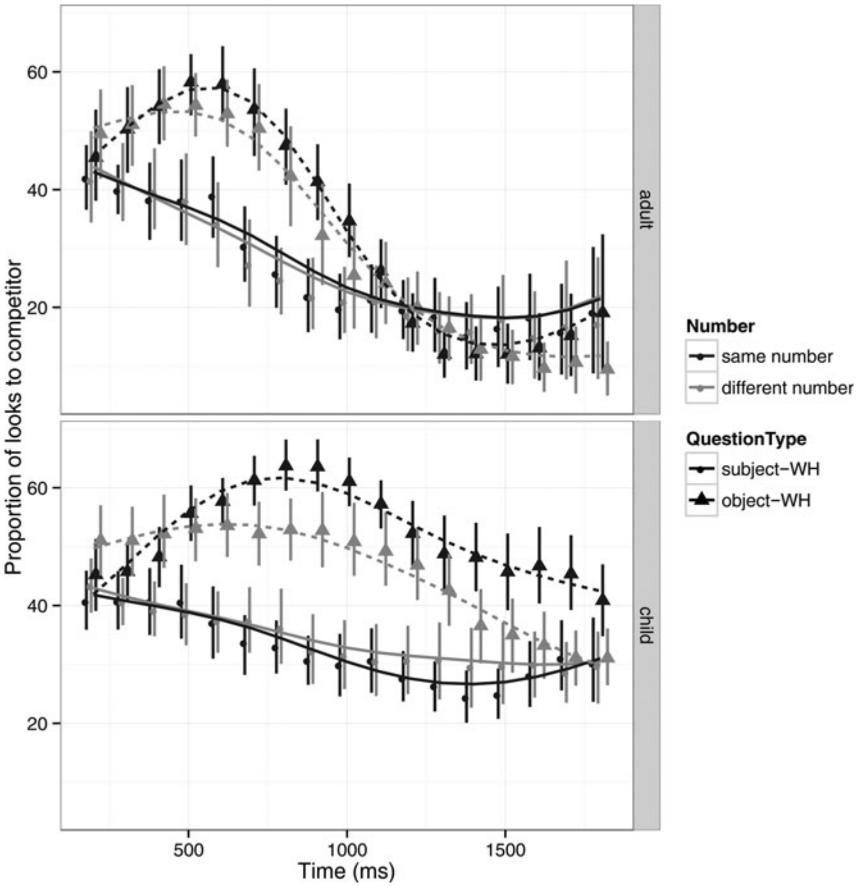

Fig. 3. Proportion of looks to the Competitor, by Group, Question type, and Number. Points show empirical means, and pointrange bars show 95% confidence intervals. Lines show model estimates.

intercepts were included, as well as all random slopes and interactions (excepting Group, which was a between-subjects factor), but no random effects correlations were included. By-Item random effects could not be included because the proportions of looks were aggregated across conditions, for each participant.

The fixed effects for both models and significance are shown in the 'Appendix'. Since our primary aim is to understand how the experimental factors and Group impact the timecourse of processing, we focus mainly on the interactions of these variables with the three Time components. To unpack these interactions, we rely on the growth curves and confidence intervals shown in Figures 2 and 3. The significant three-way interaction of Time with Group and Question type indicates differences in the shape of the growth curves for adults and children, and between *wh*-subject and *wh*-object questions. For *which*-subject questions, looks to Target increase steadily to a peak around 1100–1200 ms for both groups, with the inverse pattern apparent for looks to Competitor. Thus, children and adults appear to process *wh*-subject questions similarly.

For *which*-object questions, on the other hand, both groups show an initial decrease in looks to Target (and increase in looks to Competitor), after which gaze returns increasingly to the Target. That is, for *which*-object questions both groups initially orient to the Competitor before re-orienting to the Target. The timecourse of this sequence, however, is slower for children than for adults. Children's maximal orientation to the Competitor occurs about 200 ms later than adults', and while adults orient maximally to the Target around 1400 ms after onset of the auxiliary verb, children's looks to the Target do not reach their peak before the end of the analysis window at 1900 ms. Moreover, the trajectory of children's recovery, i.e. the increase in looks to the Target, does not level off before the end of the analysis window, as it does for the *which*-subject questions, and for the adults for both question types. Thus, while both age groups exhibit similar difficulty with *which*-object questions, the impact of this difficulty is greater for children, as seen in the slower unfolding of the growth curve.

The significant two-way interaction between Question type and Number reflects that the difficulty associated with processing *which*-object questions is greater when the subject and object match in number. However, only a marginal three-way interaction of these variables with Time was found (see 'Appendix'), and only in looks to Target ($p = \cdot 07$), yielding little evidence for an effect of number on the shapes of the growth curves in the various conditions. Moreover, the marginal three-way interaction of Question type, Number, and Group, in looks to Target ($p = \cdot 06$), provides only weak evidence for an age-based difference in sensitivity to the number manipulation for *which*-object questions. The four-way interaction of Question type, Number, Group, and Time was not significant in either model.

We also explored the contribution of child age by fitting models to the children's Target and Competitor looks only, with age in months (z-scored) as a fixed factor, with all interactions. No significant effects or interactions were found (all ps $> \cdot 07$).

DISCUSSION

This study aimed at: (1) providing detailed insights into the early stage of processing for *which*-questions in children compared to adults; and (2) testing the costs of revision during the interpretation of *which*-object questions by providing early and late cues to sentence interpretation. Accuracy analyses tested the children's off-line comprehension of subject and object *which*-questions and eye-movement analyses measured how children and adults process subject and object *which*-questions on-line.

The off-line data showed two main results: (1) children were significantly more accurate in comprehending subject questions compared to object questions; and (2) children performed more accurately in object questions with a mismatch in the Number feature between the first NP and the auxiliary, that is, structures that provide an early cue for correct interpretation. These results are in line with previous studies showing a subject–object asymmetry in English-speaking children (e.g. Avrutin, 2000, for English), and confirm previous cross-linguistic results on the comprehension of these structures (e.g. Friedmann *et al.*, 2009, for Hebrew; De Vincenzi *et al.*, 1999, for Italian). The adult data were so close to ceiling as to preclude statistical analysis, but the fact that perfect accuracy fell within the 95% CIs for all conditions except for *which*-object questions with singular subjects and objects suggests that even adults may experience some difficulty in processing these questions. Furthermore, the off-line results showed that children benefit from the earlier cue to the correct interpretation provided by number agreement on the auxiliary verb, when the *which*-phrase and auxiliary differ in number (see Contemori & Marinis, 2014, and Adani, Forgiarini, Guasti & van der Lely, 2014, for similar results on relative clauses). In addition, we did not find a significant effect of age, indicating that children aged 5;0–7;10 experience similar difficulties with object questions, in particular when the *which*-phrase and the auxiliary are both singular.

The eye-tracking data revealed a similar picture, but provided more evidence about the timecourse of children's re-orientation to the object interpretation in *which*-object questions. In the early time-windows both groups looked significantly more at the Competitor picture when an object question was presented compared to a subject question. The opposite pattern was observed for subject questions, with both adults and children looking increasingly at the Target picture until around 1100–1200 ms after the onset of the auxiliary verb. This indicates that both adults and children have a preference for a subject interpretation of the first NP (*cow*) for all question types.

However, the three-way interaction of Time × Group × Question type shows that the contrast between the shapes of the growth curves for *which*-subject and *which*-object questions was different for adults and children. That is, while the growth curves for *which*-subject questions were quite similar for adults and children, the curves for *which*-object questions differed. While both groups initially are drawn to the

Competitor and away from the Target, indicating that they try to complete the dependency at the earliest, subject gap position, adults override this bias quickly and reach maximum focus on the Target only about 200 ms later than they do for *which*-subject questions (cf. the solid curves / round points in the upper panels of Figures 2 and 3). Children, too, show evidence of recovery, but their looks to the Target do not peak within the time-window analyzed. This is in line with their accuracy results, suggesting that children are less able to overcome the difficulty of processing *which*-object questions before settling on an interpretation of the sentence, but the eye-tracking record supplies additional evidence for a developing ability to recover from their earlier commitment to the incorrect interpretation.

One possibility is that the inclusion of trials with incorrect responses in our analysis could have masked children's recovery, particularly on object questions where the number of the subject and object was the same.[2] That is, we might expect children to recover more effectively some of the time, and that when they do, they would be more likely to answer the comprehension question correctly. To explore this question, we compared the eye-gaze record for object questions in the same number condition as a function of the accuracy of children's answers (errors in the other conditions were infrequent enough, as reported above, as to preclude comparison). Visual inspection, followed up by a mixed-effects regression testing the interaction of Time and Accuracy, restricted only to this condition, revealed marginally more looks to Target overall when the comprehension question was subsequently answered correctly than when it was answered incorrectly ($p = \cdot 05$). However, the lack of any interaction with the Time components in the correct trials ($p > .95$) yielded no evidence that correct answers were preceded by more effective recovery from the initial commitment to a subject interpretation, nor of any peak in looks to Target within the time-window analyzed.

One interesting feature of these results is that the initial focus on the competitor for *which*-object questions suggests that the difficulty in processing these questions stems from a strong initial commitment, or perhaps better stated, a strong expectation that the first NP encountered will be the subject. This expectation is reflected early in the eye-movement record. Thus, it is not that children have trouble discerning the structure of the question, nor does it seem to be that their expectation of a *which*-subject interpretation is particularly strong compared to adults (they do not orient to the Target any faster for *which*-subject questions), but that they are slower in re-orienting when their initial expectation turns out to be incorrect.

[2] We thank an anonymous reviewer for pointing out this possibility.

These results are in line with previous findings on adults' processing that tested the on-line comprehension of object relative clauses in adults using eye-tracking during reading (Staub, 2010). According to Staub, at least part of the processing difficulty associated with object relatives occurs because the structure is less frequent compared to subject relative clauses. The effect of frequency was observed by Staub in the increased processing cost associated with the first encounter of the relative clause (i.e. the subject NP within the relative clause), indicating that because the parser was expecting a subject gap, the violation of this expectation resulted in a general processing difficulty with the object relative. In the present study we analyzed the early processing of *which*-questions, that are structurally similar to relative clauses, showing a similar effect to Staub. As clearly indicated by our results, the children recover more slowly from their strong expectation that the first NP refers to the subject of the sentence, as compared to the adults. We speculate that the sluggish recovery observed in the child group may sometimes affect the accuracy in the comprehension of the subject questions compared to object questions.

This result is compatible with probabilistic accounts on the processing of object dependencies (e.g. MacWhinney & Bates, 1989; Roland *et al.*, 2007; Levy, 2008), showing that, in *which*-object questions, children expect a more frequent subject-extracted structure and, when this expectation is not met, they incur a processing difficulty. The results can be interpreted within MacWhinney & Bates' (1989) Competition Model, which predicts that processing is based on the exploitation of probabilistic cue knowledge in a language. In our experiment, the strong cue to processing for English is the fact that a preverbal animate NP is reliably the agent in a sentence. According to this hypothesis, children exploit the cue and are biased to interpret that the first NP in a *which*-question as the agent.

Furthermore, the eye-tracking data add important evidence on the early stages of processing a *wh*-question, showing that children take longer to revise the first-subject preference on object questions compared to adults. Interestingly, in our study the pictures presented to the participants do not depict single referents (e.g. Trueswell *et al.*, 1999), but propositional meanings, showing that when children are given time to encode the meanings in each scene before getting to the critical parts of the sentence stimuli, they can plausibly understand and represent the differences between the scenes before making a decision. However, our design has some limitations that future research using four pictures with single referents should address.

The lack of robust group differences in the effects of number similarity between the subject and the auxiliary, on the other hand, suggests that both groups are able to take advantage of the earlier cue provided by the auxiliary verb for interpreting *which*-object questions. This sensitivity plays out on different timescales, due to children's and adults' differing sensitivity to the challenge of the *which*-object questions, but the lack of interactions of Number with Group provides no evidence that the groups differ in their sensitivity to the number cue. The finding of only marginal differences (i.e. the marginal 3-way interaction of Question type, Number, and Group in the Target looks analysis) in the present experiment is not wholly unsurprising. As shown in a recent study by Lukyanenko and Fisher (2016), children aged 2;5 to 3;0 can use verb number agreement to make predictions about the number features of upcoming nouns. Therefore, by ages 5;0–7;10 it is expected that children have considerable experience using verb number agreement to interpret complex syntactic structures, such as *which*-questions.

The structure of the stimulus set does leave one alternative interpretation of the number effects. In the number-mismatched conditions, the subject was always the plural entity, and objects were never plural. If there is a bias to interpreting plural entities as subjects, then participants' gaze might have been drawn to the correct Target picture earlier than when both subject and object were singular. There is also a possibility that participants became sensitive to the fact that plural entities were always subjects over the course of the task. Given the length of the task and the division of the experiment into two sessions, we think the latter is somewhat unlikely, and the fact that there were no significant differences across conditions in the baseline proportions of looks to Target and Competitor (nor even across age groups) at the 200 ms window suggests that this is not the case. When the child parser encounters additional information compatible with the first subject interpretation (e.g. number agreement on the auxiliary), it may be less flexible in revising it, in line with previous studies on filler–gap (Love, 1997; Omaki *et al*., 2014), garden-path sentences (Trueswell *et al*., 1999; Woodard *et al*., 2016; but see Bavin, Kidd, Prendergast & Baker, 2016, for evidence of successful revision in garden paths in five-year-old children), and passives (Huang *et al*., 2013; Marinis & Saddy, 2013).

This hypothesis also seems to match the observed off-line accuracy results, showing significantly higher accuracy for O-WH-SP compared to O-WH-SS, where no number cue is available for revision. Interestingly, the behavioral responses seem to support the hypothesis that the child parser is less flexible than the adult parser. Based on our interpretation of the results, we can speculate that the late acquisition of the object *which*-questions in English-speaking children (e.g. Avrutin, 2000) might be due to immature sentence revision mechanisms in children, as already observed by Huang *et al*. (2013) for passives.

Children's off-line accuracy results are also in line with cue-based memory retrieval models (e.g. Gordon *et al*., 2004; Van Dyke & McElree, 2006), showing that number dissimilarity between the two NPs in *which*-questions influence child off-line interpretation. For the eye-movement results, cue-based memory retrieval models predict that interference effects should be localized at the verb region where the filler-retrieval occurs. According to these predictions, if children are more affected by similarity-based interference, they should experience a difficulty (i.e. a delay in interpretation) at the end of the sentence, where the lexical verb occurs (e.g. which goat in the cow is *pushing* _). In our eye-tracking data, however, the children's difficulty is localized at the point of disambiguation, rather than at the gap. We speculate that, due to the nature of our experimental design, the subject bias plays a strong role in the commitment and revision process, and could possibly mask later interference effects. Further research using a different design is needed to address this open question.

To sum up, our data on the on-line processing of *which*-questions in children indicate that children aged between 5;0 and 7;10 demonstrate similar processing reflexes to those observed in adults. Children rely on syntactic structure in their on-line sentence processing comparably to adults, and can successfully interpret also the harder *which*-object questions. When information disambiguating for a subject/object interpretation is provided early, our results suggest that children may be more likely to interpret the *which*-object questions correctly. On the other hand, when children have already committed to one of the interpretations, their reanalysis may be less efficient compared to adults, and children are more likely to entertain the first (subject) interpretation. We speculate that cognitive constraints could play a role in children's reduced flexibility in reanalysis. It is well known that a number of cognitive mechanisms that interact with language processes, such as working memory or cognitive control, undergo substantial development during language development (e.g. Mendelsohn, 2002; Novick *et al.*, 2010; Woodard *et al.*, 2016). For instance, Mendelsohn (2002) has found that the size of participants' garden-path effect correlates with several linguistic and non-linguistic measures all of which involve inhibition/selection mechanisms. Furthermore, we do not exclude the possibility that additional factors, such as the ability to detect the statistical distribution of the input, may contribute to the late development of revision capabilities (e.g. Kidd & Arciuli, 2016). For instance, Kidd and Arciuli showed that individual differences in Statistical Learning (SL, i.e. the ability to attend to statistical regularities in the input) predict the comprehension performance on object relative clauses in six- to eight-year-old children. However, in our study we did not measure the role of EF or SL, hence we cannot exclude other interpretations for the revision difficulty observed in children's interpretation. Therefore, future research should investigate the relationship between EF, SL, and the processing and off-line interpretation of *which*-object questions. Additionally, because our study does not provide strong evidence for a clear role of number as an early revision cue in child processing of *which*-questions, future studies should address this issue further.

CONCLUSIONS

The results of the present study show a bias for an agent interpretation of the first NP in object questions in adults and children. Our results provide details on the timing of the dependency completion, showing that children show less flexibility in the process of reanalysis for object questions compared to subject questions. These results provide support for the hypothesis that children's lower accuracy in the comprehension of object questions is due at least partly to difficulties in processing, and in particular with the process of revising an agent-interpretation for the first NP in the sentence. Our study also suggests that the number information on the auxiliary verb in object questions can be an effective syntactic cue that guides adults' and children's processing of filler–gap dependencies.

REFERENCES

Adani, F., Forgiarini, M., Guasti, M. T. & van der Lely, H. K. J. (2014). Number dissimilarities facilitate the comprehension of relative clause in children affected by (Grammatical) Specific Language Impairment. *Journal of Child Language* **41**(4), 811–41.

Agresti, A. (2002). *Categorical data analysis*, 2nd ed. New York: John Wiley & Sons.

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In Yosef Grodzinsky, Lee Shapiro & David Swinney (eds). *Language and brain: representation and processing* (pp. 295–313). San Diego, CA: Academic Press.

Baayen, R. H., Piepenbrock, R. & Gulikers, L. (1995). *The CD-ROM version of the CELEX lexical database*. Nijmegen: Max Planck Institute for Psycholinguistics.

Bates, D., Maechler, D., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**(1), 1–48.

Barr, D. J. (2008). Analyzing 'visual world' eye-tracking data using multi-level logistic regression. *Journal of Memory and Language* **59**(4), 457–74.

Bavin, E., Kidd, E., Prendergast, L. A. & Baker, E. K (2016). Young children with ASD use lexical and referential information during on-line sentence processing. *Frontiers in Psychology* **7**, 171. Online: <doi: 10.3389/fpsyg.2016.00171>.

Bever, T. G. (1970). Cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the development of language*, 279–362. New York: Wiley.

Christianson, K., Hollingworth, A., Halliwell, J. F. & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology* **42**, 368–407.

Contemori, C. & Marinis, T. (2014). The impact of number mismatch and passives on the real-time processing of relative clauses. *Journal of Child Language* **41**(3), 658–89.

Deevy, P. & Leonard, L. (2004). The comprehension of Wh-questions in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* **47**, 802–15.

De Vincenzi, M., Arduino, L. S., Ciccarelli, L. & Job, R. (1999). Parsing strategies in children: comprehension of interrogative sentences. In S. Bagnara (Ed.), *Proceedings of*

European Conference on Cognitive Science (pp. 301–308). Rome: Instituto di Psicologia del CNR.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.

Diessel, H. (2009). *The emergence of relative clauses in early child language*. Unpublished ms, University of Jena, Thuringia, Germany.

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

Frazier, L. & Flores d'Arcais, G. B. (1989). Filler-driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language* **28**, 331–44.

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

Goodluck, H. (2005). D(iscourse)-linking and question formation: comprehension effects in children and Broca's aphasics. In A. M. Di Sciullo (ed.), *UG and external systems: language, brain and computation* (pp. 185–192). Amsterdam: John Benjamins Publishing Company.

Gordon, P., Hendrick, R. & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* **51**, 97–114.

Guasti, M. T. (1996). The acquisition of Italian interrogatives. In H. Clahsen (ed.), *Generative perspectives on language acquisition* (pp. 241–269). Amsterdam: John Benjamins.

Harrell, F. (2001). *Regression modeling strategies*. New York: Springer.

Huang, Y., Zheng, X., Meng, X. & Snedeker, J. (2013). Assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language* **64** (4), 589–606.

Jakubowicz, C. & Gutierrez, J. (2007). *Elicited production and comprehension of root wh-questions in French and Basque*. Presentation at the COST Meeting Cross-linguistically robust stage of children's linguistic performance, Berlin.

Kidd, E. & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, **87**(1), 184–193.

Kidd, E., Brandt, S., Lieven, E. & Tomasello, M. (2007). Object relatives made easy: a crosslinguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes* **22**, 860–97.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* **106**, 1126–77.

Love, T. (1997). The processing of non-canonically ordered constituents in long distance dependencies by pre-school children: a real-time investigation. *Journal of Psycholinguistic Research* **36**, 191–206.

Lukyanenko, C. & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition* **146**, 349–70.

MacWhinney, B. & Bates, E. (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Marinis, T. & Saddy, D. (2013). Parsing the passive: comparing children with Specific Language Impairment to sequential bilingual children. *Language Acquisition* **20**, 155–79.

Matin, E., Shao, K. & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perception & Psychophysics* **53**, 372–80.

Mendelsohn, A. (2002). *Individual differences in ambiguity resolution: working memory and inhibition*. Unpublished PhD dissertation, Department of Psychology, Northeastern University, Boston, MA.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press.

Novick, J. M., Kan, I. P., Trueswell, J. C. & Thompson-Schill, S. (2010). A case for conflict across multiple domains: memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology* **26**(6), 527–67.

Novick, J. M., Trueswell, J. C., Thompson-Schill, S. L. (2005). Executive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective & Behavioral Neuroscience* **5**, 263–81.

Omaki, A., Davidson-White, I., Goro, T., Lidz, J. & Phillips, C. (2014). No fear of commitment: children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development* **10**(3), 206–33.

Roland, D., Dick, F. & Elman, J. L. (2007). Frequency of English grammatical structures: a corpus analysis. *Journal of Memory and Language* **57**, 348–79.

Schlesewsky, M., Fanselow, G., Kliegl, R. & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (eds), *German sentence processing*, 65–93. Dordrecht: Kluwer.

Schneider, W., Eschman, A. & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: University of Pittsburgh Psychology Software Tools, Inc.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition* **116**, 71–86.

Stavrakaki, S. (2006). Developmental perspectives on specific language impairment: evidence for the production of wh-questions by Greek SLI children over time. *Advances in Speech–Language Pathology* **8**, 384–96.

Stowe, A. (1986). Parsing wh-constructions: evidence for on-line gap location. *Language and Cognitive Processes* **1**, 227–45.

Stromswold, K. (1995). The cognitive and neural bases of language acquisition. In M. Gazzaniga (ed.), *The cognitive neurosciences*, 855–70. Cambridge, MA: MIT Press.

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**, 1632–1634.

Trueswell, J. & Gleitman, L. R. (2004). Children's eye movements during listening: evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (eds). *Interface of language, vision, and action: eye movements and the visual world* (pp. 319–346). New York: Psychology Press.

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

Van Dyke, J. & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language* **55**, 157–66.

Wilson, M. (1988). MRC psycholinguistic database: machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, and Computers* **20**, 6–10.

Woodard, K., Pozzan, L. & Trueswell, J. (2016). Taking your own path: individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology* **141**, 187–209.

Yoshinaga, N. (1996). *Wh-questions: a comparative study of their form and acquisition in English and Japanese*. Unpublished doctoral dissertation, University of Hawaii.

APPENDIX: fixed effects for models of Target and of Competitor looks

	Target					Competitor				
Dependent variable:	β	SE	χ^2	*df*	*p*	β	SE	χ^2	*df*	*p*
Intercept	0·17	0·04				−0·67	0·04			
Time_1	0·13	0·02	83·59	3	<·0001	−0·24	0·02	113·99	3	<·0001
Time_2	−0·41	0·06				0·28	0·05			
Time_3	1·54	0·32				−2·13	0·32			
Question Type	−0·44	0·05	46·48	1	<·0001	0·50	0·06	43·09	1	<·0001
Number	0·12	0·04	6·95	1	·008	−0·10	0·05	3·66	1	·06
Group	−0·51	0·08	28·82	1	<·0001	0·66	0·08	42·14	1	<·0001
Time_1 × Question Type	0·16	0·03	59·13	3	<·0001	−0·14	0·04	39·22	3	<·0001
Time_2 × Question Type	0·30	0·10				−0·35	0·11			
Time_3 × Question Type	3·76	0·68				−3·26	0·73			
Time_1 × Number	0·04	0·03	6·37	3	·09	−0·05	0·04	6·50	3	·09
Time_2 × Number	−0·03	0·11				0·08	0·11			
Time_3 × Number	−1·23	0·64				1·37	0·64			
Question Type × Number	0·31	0·09	9·89	1	·002	−0·24	0·10	6·03	1	·01
Time_1 × Group	−0·12	0·04	39·66	3	<·0001	0·26	0·05	56·00	3	<·0001
Time_2 × Group	0·64	0·12				−0·59	0·11			
Time_3 × Group	−2·17	0·66				1·87	0·66			
Question Type × Group	−0·39	0·11	12·30	1	·0004	0·48	0·13	12·96	1	·0003
Number × Group	0·04	0·09	0·20	1	·65	−0·01	0·11	0·01	1	·94
Time_1 × Question Type × Number	0·08	0·05	7·06	3	·07	−0·09	0·07	5·34	3	·15
Time_2 × Question Type × Number	−0·29	0·20				0·22	0·20			
Time_3 × Question Type × Number	−1·68	1·18				1·76	1·18			
Time_1 × Question Type × Group	−0·12	0·06	18·42	3	·0003	0·22	0·08	20·52	3	·0001
Time_2 × Question Type × Group	0·62	0·20				−0·61	0·22			
Time_3 × Question Type × Group	−3·25	1·39				3·49	1·49			
Time_1 × Number × Group	0·02	0·05	0·93	3	·82	−0·02	0·08	0·35	3	·95
Time_2 × Number × Group	−0·15	0·22				0·12	0·23			
Time_3 × Number × Group	0·68	1·30				−0·01	1·32			
Question Type × Number × Group	0·36	0·19	3·41	1	·06	−0·25	0·20	1·54	1	·22
Time_1 × Question Type × Number × Group	0·10	0·10	4·23	3	·24	−0·05	0·15	2·03	3	·57
Time_2 × Question Type × Number × Group	−0·61	0·40				0·53	0·42			
Time_3 × Question Type × Number × Group	2·27	2·40				−0·92	2·42			