Children's eye gaze reveals their use of discourse context in object pronoun resolution

Jacolien van Rij[1], Bart Hollebrandse[2], Petra Hendriks[2]
[1]Eberhard Karls University, Tübingen, Germany
[2]University of Groningen, The Netherlands

Abstract

Dutch-speaking children, like English-speaking children, allow a non-coreferential as well as a coreferential interpretation for an object pronoun. Our question is whether children make use of the structure of the discourse context and the prominence of referents to resolve this pronominal ambiguity. We describe a visual world experiment with Dutch children (4-6 years old) and a control group of adult participants. Three different context types were tested, which varied in the order of referent introduction and the number of referents that were introduced. The children's behavioral responses suggest that their performance becomes more adult-like when the antecedent is more prominent in the context. The gaze data shows that visual and linguistic context can direct children towards an adult-like interpretation, but also towards an ungrammatical interpretation. This suggests that children between 4-6 years old rely less on grammatical knowledge to resolve an object pronoun than adults, but also use other information such as discourse prominence and visual context.

Keywords: discourse processing, language acquisition, language processing, object pronouns, visual world paradigm

Introduction

In languages such as English and Dutch, the interpretation of pronouns in object position (e.g., *him* or *her*) is constrained by the structure of the sentence. An object pronoun cannot refer to the subject of the same clause, but has to refer to another referent

Contact: vanrij.jacolien@gmail.com. The authors would like to thank Robbert Prins and Petra van Berkum for drawing the pictures for this study, and Laura Hemstra for recording the sound files. Furthermore, the audience of the GLOW workshop *The timing of grammar* in Potsdam in 2012 and the reviewers of this paper are thanked for their useful comments. This investigation was supported by a grant from the Netherlands Organization for Scientific Research (NWO, grant no. 277-70-005). Data and R scripts for all analyses are available upon request and will be made available on . . . , `www.website-of-book.com`.

in the sentence or in the discourse. For example, in a context with a squirrel and a rabbit the object pronoun *him* in Sentence (1) can only refer to the rabbit, but not to the squirrel (cf. Principle B of Binding Theory; Chomsky, 1981). When a speaker would have wished to describe the situation that the squirrel was pointing to himself, he would have used a reflexive (e.g. *himself* or *herself*). When used in object position, pronouns and reflexives are therefore in complementary distribution (e.g. Chomsky, 1981).

(1). The squirrel is pointing at him with a stick.

Many studies have shown that children up to the age of 6 resolve object pronouns (e.g., *him* or *her*) differently from adults in languages such as English and Dutch (e.g., Chien & Wexler, 1990; Koster, 1993; Matthews, Lieven, Theakston, & Tomasello, 2009; Philip & Coopmans, 1996). Whereas adults only allow a non-coreferential interpretation for sentences such as (1) (i.e., the squirrel is pointing to the rabbit), children also allow a coreferential interpretation for the object pronoun (i.e., the squirrel is pointing to himself). The same children show adult-like interpretation of reflexives much earlier in their acquisition, from around 4 years old on (e.g., Chien & Wexler, 1990; Koster, 1993). Recent studies have found that children's production of pronouns and reflexives in object position is also adult-like from at least 4 years old on (De Villiers, Cahillane, & Altreuter, 2006; Matthews et al., 2009; Spenader, Smits, & Hendriks, 2009).

Different accounts have been put forward to explain this specific delay in the acquisition of object pronoun comprehension. Within the tradition of generative syntax, it has been proposed that children have the grammatical knowledge that object pronouns should not be bound in the local clause (cf. Binding Theory, Chomsky, 1981), but lack the pragmatic knowledge to restrict the co-referential interpretation to the few specific contexts in which it is allowed (e.g., Chien & Wexler, 1990; Grodzinsky & Reinhart, 1993; Thornton & Wexler, 1999). On the other hand, within the Optimality Theory framework (Prince & Smolensky, 2004), Hendriks (2014) argues that the grammar itself is asymmetrical and allows for a co-referential interpretation. Children need to learn to use perspective taking (i.e., as a listener learn to consider the linguistic choices of the speaker, formalized in Optimality Theory as bidirectional optimization) to infer that the speaker would have used a reflexive rather than a pronoun when he/she wished to express a co-referential interpretation. Her theory does not only account for the differences between pronoun comprehension and reflexive comprehension, but also provides an explanation for the difference between pronoun comprehension and pronoun production. In contrast to the accounts mentioned above, usage-based accounts of pronoun acquisition argue that children have not yet fully acquired the grammatical knowledge to reject the co-referential interpretation. These accounts assume that this knowledge is only gradually learned from the lexical input (e.g., Matthews et al., 2009).

In addition to grammatical knowledge, also processing factors may play a role. Van Rij, van Rijn, and Hendriks (2010) found that children's performance on object pronoun comprehension improves when they have more time for interpretation due to slower speech rate. This suggests that when children have acquired the relevant grammatical knowledge, they may not yet posses sufficient processing speed to complete the process of perspective taking required to arrive at the adult interpretation.

Assuming that pronouns are ambiguous for children, as Hendriks (2014) proposes, children may rely on other sources of information to resolve pronouns, such as the prominence of referents in the discourse context. Indeed, Spenader et al. (2009) report that with simple transitive sentences such as (1) children between 4 and 6 years old correctly reject the co-referential interpretation more often when only the patient (e.g., *the rabbit*) is introduced linguistically, rather than both referents. Introducing only the patient makes this referent more prominent in the discourse than when both referents are introduced. Within the tradition of generative syntax, on the other hand, pronouns are not considered to be ambiguous for children, because it is assumed that children use the same grammatical principles as adults (e.g., Chien & Wexler, 1990). Because Spenader et al. (2009) only measured behavioral responses, their experiment does not provide information on how context influences the *on-line* processing of object pronouns. In the current study, we aim to investigate how context influences children's object pronoun processing by measuring their eye gaze.

Children's eye movements during pronoun processing

Different studies using the visual world paradigm (e.g. Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) have shown that children as young as 2–4 years old are sensitive to context information, showing a preference for looking at the firstly introduced referent or subject referent when hearing an ambiguous *subject* pronoun such as *he* or *she* (e.g., Järvikivi, Pyykkönen-Klauck, Schimke, Colonna, & Hemforth, 2014; Pyykkönen, Matthews, & Järvikivi, 2010; Song & Fisher, 2005, 2007), similar to adults (e.g., Gernsbacher & Hargreaves, 1988; Gordon, Grosz, & Gilliom, 1993; Järvikivi, van Gompel, Hyönä, & Bertram, 2005; Kaiser & Trueswell, 2008). These effects occur relatively late, not within 1000 ms (e.g., Arnold, Brown-Schmidt, & Trueswell, 2007).

The eye-tracking studies that investigated the acquisition of *object* pronouns show mixed results. Sekerina, Stromswold, and Hestvik (2004) investigated children's and adults' looking behavior while listening to sentences with a pronoun or reflexive in a prepositional phrase (e.g., The boy has put the box behind him/himself.). Note that pronouns in prepositional phrases are grammatically ambiguous for adults, in contrast with pronouns in object position (cf. Runner, Sussman, & Tanenhaus, 2003). They found that after hearing the pronoun, children and adults look more to the picture with a co-referential interpretation than to the picture with a non-coreferential interpretation. Children's preference for the co-referential picture started around 1000 ms later than adults' preference.

To investigate pronouns in object position, which are unambiguous for adults, Bergmann, Paulus, and Fikkert (2011) performed an eye-tracking study and a picture selection task with 3- and 4-year-old Dutch children. Analyzing the gaze patterns from 200–700 ms after onset of the referring expression, they found that 4-year-olds, but not 3-year-olds, look more to the correct antecedent of the pronoun or reflexive than in a baseline condition where no reference was made to that character. On the other hand, the children did not show adult-like performance on the picture selection task, but preferred a co-referential interpretation for pronouns. Bergmann et al. (2011) conclude that children from 4 years old on have the grammatical knowledge necessary for pronoun comprehension, and that their observed difficulties in pronoun interpretation may arise from task effects (cf. Conroy, Takahashi, Lidz, & Phillips, 2009). However, children's performance on the gaze data and the picture selection task are difficult to compare in this study, because in the eye-tracking task only pictures with a grammatical interpretation were presented (i.e., showing an other-oriented action for a pronoun sentence and a self-oriented action for a reflexive sentence). Therefore, it is not clear whether the 4-years-olds would show similar gaze behavior when presented with a co-referential interpretation while listening to a pronoun sentence.

Clackson, Felser, and Clahsen (2011) recorded the eye gazes of 6- to 9-year-old English-speaking children on sentences with two referents. In four different regions of the screen, the antecedent referent, a highly prominent ungrammatical referent, a referent mentioned as an object in the sentence, and a distractor referent were presented. The results indicate that children between 6 and 9 years old direct their gaze to the grammatical referent shortly after pronoun or reflexive onset. However, when the highly prominent ungrammatical referent matched in gender with the pronoun, they also looked to that referent. The same children showed adult-like performance on a picture verification task. Adults are also distracted by the ungrammatical referent when hearing a pronoun, but less so when hearing a reflexive. Clackson et al. (2011) argue that discourse context and grammatical principles interact during pronoun processing (cf. Badecker & Straub, 2002, but see Nicol & Swinney, 1989).

In contrast with the conclusion of Bergmann et al. (2011), the study of Clackson et al. (2011) shows that children look more to the ungrammatical referents than adults, although their behavioral responses indicate adult-like performance. Interestingly, in the studies of Bergmann et al. (2011) and Clackson et al. (2011) children were not much slower than adults in directing their looks to the grammatical antecedent (around 400 ms after pronoun onset), whereas Sekerina et al. (2004) report much later effects for children than for adults.

Thus, previous eye-tracking studies on the acquisition of object pronouns yielded different conclusions on whether children posses adult-like grammatical knowledge for object pronoun interpretation (e.g., Bergmann et al., 2011; Clackson et al., 2011). Eye-tracking studies also showed inconclusive results with respect to the timing of processing of ambiguous pronouns. Some studies report later effects for children than for adults (e.g., Arnold et al., 2007; Järvikivi et al., 2014; Sekerina et al., 2004), whereas other studies show adult-like timing for children (e.g., Bergmann et al., 2011; Clackson et al., 2011). Context seems to influence children's object pronoun processing (e.g., Clackson et al., 2011; Spenader et al., 2009), but it is not clear what discourse factors facilitate or hinder children's object pronoun processing.

Current study

Our study investigates whether and how children who do not yet interpret object pronouns in an adult-like way make use of context to resolve the pronoun. In line with Spenader et al. (2009) for children and Hendriks, Banga, van Rij, Cannizzaro, and Hoeks (2011) for adults, we contrasted a single-referent context with a two-referent context to manipulate the prominence of the antecedent in the discourse. In addition, we manipulated the order of referent introduction within the two-referent contexts to investigate the effect of discourse structure further. Examples of the three different context manipulations for Example (1) are listed in (2).

(2). Context manipulations:

Two-referent contexts:	Agent-Patient	**(AP)**:	Hier zie je een eekhoorn en een konijn. 'Here you see a squirrel and a rabbit.'
	Patient-Agent	**(PA)**:	Hier zie je een konijn en een eekhoorn. 'Here you see a rabbit and a squirrel.'
Single-referent context:	Patient	**(P)**:	Hier zie je een konijn. 'Here you see a rabbit.'
Test sentence:	The eekhoorn wijst naar hem / zichzelf met een stok. 'The squirrel is pointing at him / himself with a stick.'		

The reflexive sentences functioned as a control condition, because children generally acquire adult-like performance on reflexive interpretation before they acquire adult-like performance on object pronoun interpretation (cf., Başkent, van Rij, Ng, Free, & Hendriks, 2013; Van Rij et al., 2010). Based on an earlier study comparing the behavioral responses in a single-referent context with a two-referent context (Spenader et al., 2009), we expect more adult-like performance on children's pronoun comprehension with the single-referent context than with the two-referent contexts, because the antecedent is highly prominent in that condition. In the gaze data we expect that the single-referent context results in more looks to the antecedent after hearing the pronoun than the two-referent contexts. When discourse prominence plays a role, we would expect that children's performance on pronoun sentences also increases with a PA context in comparison with an AP context. In the PA context, the pronoun antecedent is more prominent than in the AP context, because it is introduced in first position (cf. Gernsbacher & Hargreaves, 1988; Kaiser & Trueswell, 2008).

Methods

Participants

Sixty-one monolingual Dutch children (4-6 years old) were tested in the Eye Lab of the University of Groningen, The Netherlands. The data of twenty children was excluded from analysis: 1 child was using medication, 5 children could not finish the task because of technical problems, and 14 children answered less than 75% of the control trials correctly, suggesting that they did not understand the task. Some of these children gave world knowledge answers, such as 'bears do not tickle'. Thirty-nine adults were asked to participate as control group. Two adult participants could not finish the task because of technical problems, and the data of another adult participant was excluded from analysis because this participant was not a native speaker of Dutch. Adult participants were compensated with course credits for their participation, whereas children were given a small toy.

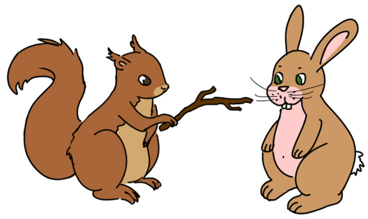
The data of 41 children (age range: 4;0–6;5, M_{age} = 5;2, 25 boys) and 36 adults (age range: 17–64, $M_{age} = 25$, 7 men) was analyzed.

Design and materials

Participants performed a picture verification task. In this task, they were asked to judge whether a pre-recorded sentence that was presented auditorily was a correct description of the picture presented on the screen.

In a 2 x 2 x 3 design we manipulated *a*) the referring expression in the test sentence, a pronoun or a reflexive; *b*) the picture presented on the screen, showing an other-oriented action (congruent with a pronoun sentence, but incongruent with a reflexive sentence) or a self-oriented action (incongruent with a pronoun sentence, but congruent with a reflexive sentence); *c*) the introduction of referents in the context sentence, AP, PA, or P. This results in twelve variants for each of the twenty-four items. The three different contexts for a pronoun sentence are illustrated in (2), and Figure 1 shows the pictures for the same item. All pictures show two referents, and one of these is performing an action with an instrument. We will refer to these referents as *agent* and *patient*, although in the self-oriented pictures (Figure 1 *b.*) the non-acting referent is actually not a patient of the action.

Six versions of the experiment were constructed by a Latin square design, with each list containing 32 test items. The following verbs were used in the test sentences: *fo-*



a. Other-oriented picture
(congruent with pronoun)

b. Self-oriented picture
(incongruent with pronoun)

Figure 1. Examples of the two types of pictures with the verb *aanwijzen* 'to point at'.

tograferen 'to take a picture', *filmen* 'to film', *tekenen* 'to draw', *schilderen* 'to paint', *schoppen* 'to kick', *slaan* 'to hit', *steken* 'to poke', *kietelen* 'to tickle', *aanwijzen* 'to point at', *aanraken* 'to touch', *optillen* 'to lift', *knijpen* 'to pinch'. Some of the referent pairs and verbs appeared twice in the list.

Each participant was tested on 32 test items, 16 of which were pronoun sentences and 16 were reflexive sentences. Each of these sets of sentences was divided in 8 congruent and 8 incongruent items. Of these subsets, 4 items were preceded by a single-referent context, and 4 by a two-referent context (2 AP and 2 PA). In addition to these 32 test items, 7 filler items without a referring expression were randomly distributed over the list to test whether children understood the task. Pictures were flipped from left to right in half of the trials (randomly determined). For one test item, incorrectly an other-oriented action was shown instead of a self-oriented action when the image was flipped. The scores on this item were corrected before the analysis.

Procedure

For children, the experiment was explained by a hand puppet. The experiment started with a calibration, followed by three practice trials (without a pronoun or reflexive) to familiarize participants with the procedure. A trial started after the eye-tracker had detected a 500 ms long fixation on the fixation cross. First the picture appeared on the screen and 1000 ms later the pre-recorded sentence started. Using a button box, participants had to press on a green happy smiley face when the sentence was congruent with the picture and on an orange sad face when the sentence was incongruent with the picture. After pressing one of the response buttons, the selected smiley was presented on the screen to confirm the button press, but no other feedback was provided on the answers. After the training trials, half-way the experiment, and at the end of the experiment general positive feedback was provided (e.g., 'Fantastic!' or 'Well done!') to keep the children motivated. The items were presented in two blocks, with a short break in between. The full experiment took around 20 minutes.

The participants' eye movements were recorded with a remote Tobii T120 eyetracker in a two computer setup. One computer presented the stimuli and recorded behavioral responses with the experiment software E-Prime 2.0 (Psychological Software Tools, Inc.). The other computer collected the gaze data, which was recorded with a sample rate of 120 Hz. The gaze positions reported are the averages of both eyes with a high tracking validity, or the gaze position based on one eye when the tracking validity of the other eye was low.

Results

The data was analyzed using Generalized Additive (Mixed) Modeling (GAMM; Lin & Zhang, 1999; Wood, 2006, 2011). GAMM is a nonlinear regression method with the possibility to include linear and nonlinear random effects. The method is particularly suited for analyzing time course data, such as gaze data (e.g., Nixon, van Rij, Mok, Baayen, & Chen, submitted), pupil dilation (e.g., van Rij, Natalya, van Rijn, Wood, & Baayen, submitted), EEG (e.g., Boehm, van Maanen, Forstmann, & van Rijn, 2014), or articulography data (e.g., Tomaschek, Wieling, Arnold, & Baayen, 2013), because GAMMs can fit nonlinear trends over time that are typical for these measures. The relation between the dependent variable and the predictor is modeled as a smooth function, which is a weighted sum of a set of base functions that each have a different shape. Together these weighted functions can fit linear and nonlinear patterns of the data. The estimation procedures determining the smooth functions and parameters are designed to avoid overgeneralization and overfitting of the data (Wood, 2006). The behavioral responses and the gaze positions were analyzed in R (version 3.1.2; R Core Team, 2014) using the packages *mgcv* 1.8.3 (Wood, 2006) and *itsadug* 0.9 (van Rij, Wieling, Baayen, & van Rijn, 2015).

Behavioral results

Figure 2, leftmost panels, shows the accuracy of the judgements by children and adult participants for the sentences with a pronoun (*him*) or a reflexive (*himself*). Adult participants gave only a few incorrect answers (9 out of 1152). Children gave more incorrect answers, especially on the incongruent items (196 for the incongruent items versus 67 for the congruent items) suggesting a 'yes'-bias (cf. Chien & Wexler, 1990). Our results replicate the findings of Spenader et al. (2009): children's pronoun comprehension increased from 69% adult-like responses (69% in their study) after a two-referent context to 77% adult-like responses (83% in their study) when only the patient was mentioned in the introduction sentence.

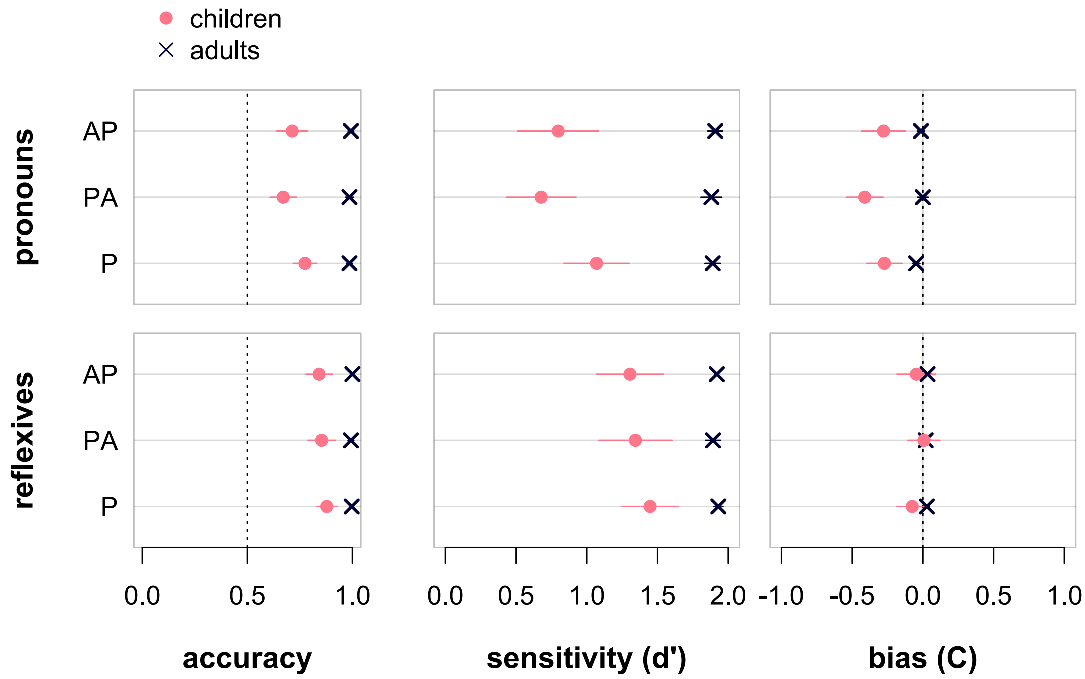


Figure 2. Dotplot of average participant means (±95%CI) of behavioral measures. *Left:* Accuracy, *Center:* Sensitivity, *Right:* Response bias.

Sensitivity and bias. The answers of the child and adult participants were converted into two measures based on Signal Detection Theory (SDT; Macmillan & Creelman, 2004; Stanislaw & Todorov, 1999), the sensitivity d' and the response bias C. The sensitivity d' reflects how well participants can distinguish between congruent and incongruent trials, with a higher (positive) value of d' indicating more correct 'yes' responses on congruent trials and fewer incorrect 'yes' responses on incongruent trials (cf. Başkent et al., 2013). The response bias C reflects the participants' bias for saying 'yes': a positive value indicates that participants are more likely to give incorrect answers on congruent items than on incongruent items ('no' bias), a negative value indicates that participants are more likely to give incorrect answers on incongruent items than on congruent items ('yes' bias), and a value around zero indicates that participants are equally likely to give incorrect answers on congruent and incongruent items (no bias). See the Supplementary Materials for a more precise description of the SDT measures.

SDT measures were used rather than accuracy, because they are able to disentangle the sensitivity from potential response biases. Figure 2, center and rightmost panels, shows the sensitivity and bias for children (red dots) and adults (black crosses). Children's sensitivity on pronoun items (top row) is lower than on reflexive items (bottom row), and lower than adults' sensitivity. In addition, children show a clear 'yes'-bias in their responses to pronoun items, but not in their responses to reflexive items.

GAMM analysis. As the d' and C measures describe different aspects of the same set of responses, we analyzed the measures in a single analysis. The measures were combined into a single dependent variable, but with different labels for each measure. Using GAMM we tested whether the structure of the context influences children's sensitivity and bias. We did not include adults' responses in the analysis, because they show ceiling performance.

All interactions between the *Measures* (d' or C), *Referring Expression* (pronouns or reflexives), and *Context* (AP, PA, or P) were combined in one predictor. Interactions between this predictor and Age were included in the model as nonlinear smooths. For each participant a random intercept was included, and random slopes for the differences between d' and C, and between pronouns and reflexives.

An iterative backward-fitting model comparison procedure was employed to find the best fitting model starting with the full model. We additionally tested the differences between conditions using custom contrasts (cf. Masson & Kliegl, 2013, paper package). The best fitting models are presented in Appendix 1. Although their structure seems different, the models actually result in similar predictions. The Supplementary Materials present all statistical analyses and visualizations of the estimated differences between the context conditions.

Inspection of the statistical models revealed a significant interaction between the participant's age (in months) and the SDT measure: the d' values increase with age, whereas the C values do not increase with age. Figure 3 shows the model's predictions for d' and C. In addition, the model revealed a significantly lower sensitivity d' on pronoun items in comparison to reflexive items (β=-.419, SE=.078, t=-5.36) and also a significantly lower bias (although smaller difference in bias, β=0.242, SE=.069, t=3.50). The *Context* conditions only influence the sensitivity d' in pronoun comprehension, but no significant differences were found in the response bias C, and marginally significant differences on the sensitivity d' in reflexive comprehension (when the smooth terms are taken into account the significant intercept differences in Table 1 disappear, see Supplementary Materials). A higher sensitivity was found for the single-referent contexts (P) than for the two-referent contexts (AP and PA) in pronoun processing. In addition, the two-referent contexts show much more variation, and a less clear trend than the single-referent context.

Thus, children show a significantly higher sensitivity for reflexives than for pronouns and a larger 'yes'-bias with pronouns than with reflexives, confirming the often found delayed acquisition of pronoun interpretation in comparison with reflexive interpretation. The difference in response bias between pronouns and reflexives indicates that the task itself was not too difficult for children, for example due to the inclusion of incongruent pictures, because then also reflexives would have shown a significant response bias. More importantly

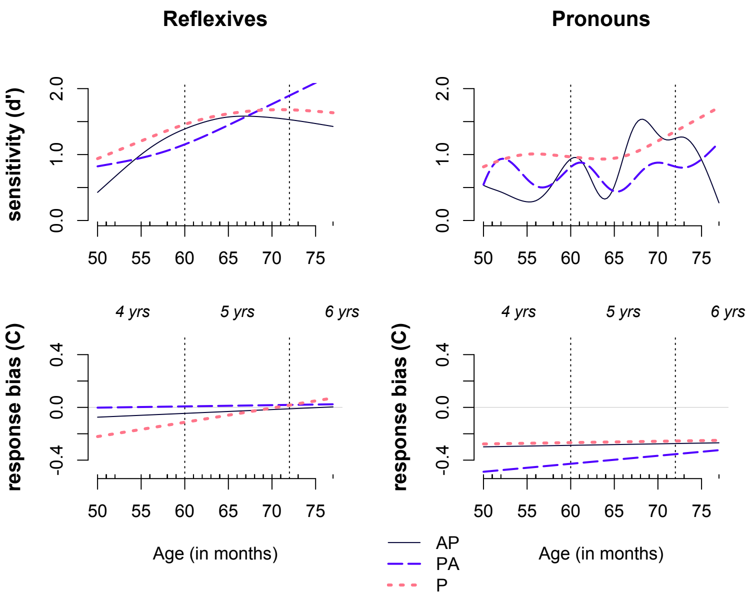


Figure 3. Estimations of the change in children's sensitivity d' (*Top*) and response bias C (*Bottom*) over Age for Reflexive items (*Left*) and Pronoun items (*Right*), after removing the random effects for individual participants. The black solid lines are the estimated effects for the AP context, the blue dashed lines are the estimates for the PA context conditions, and the pink dotted lines are the estimated effects for the P context.

for our research question is that children's sensitivity for pronouns is increased with a single-referent context, in which only the patient is introduced. This is in line with the results of Spenader et al. (2009), although in our study the single-referent condition does not improve children's pronoun comprehension to the same level as their reflexive comprehension.

Gaze data

The analysis of the gaze data focuses on the pronoun items. To investigate whether children are sensitive to the discourse structure and prominence of referents in the preceding context, we analyzed their gaze patterns following the pronoun and compared these with the gaze patterns of the adult participants. The data from 1000 ms before pronoun onset until 3000 ms after pronoun onset was selected for analysis. Each picture contained three areas of interest: the agent, the patient, and the instrument. All recorded gaze positions were mapped automatically to these regions in R (R Core Team, 2014). The looks to other areas of the screen were not considered in the analysis. The data was divided into bins of 200 ms, and within each time bin the number of looks to the agent, to the patient referent, and to the instrument were counted for each participant and for each item.

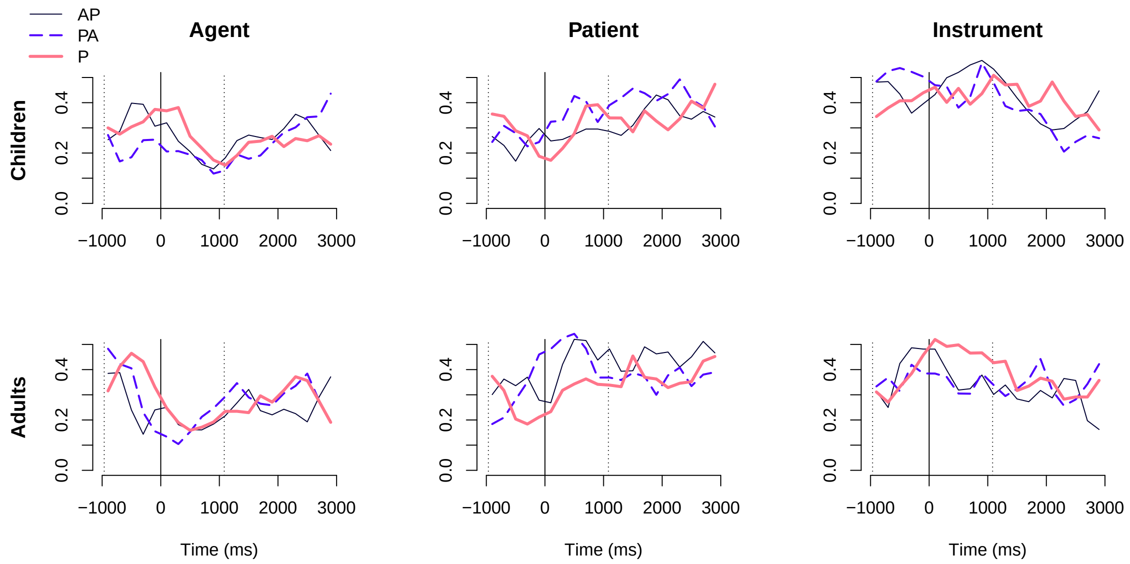


Figure 4. Gaze data when viewing a **congruent picture** with other-oriented action, for children (*top row*) and adults (*bottom row*). The plots show separate lines for the three context conditions: AP (black thin line), PA (blue dashed line), and P (pink thick line).

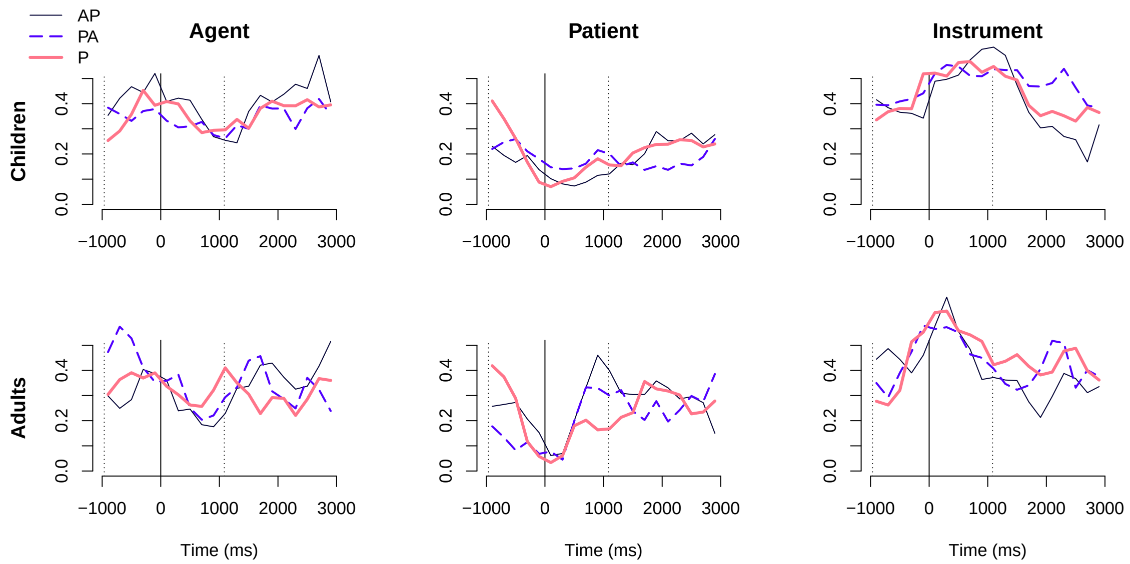


Figure 5. Gaze data when viewing an **incongruent picture** with self-oriented action, for children (*top row*) and adults (*bottom row*). The plots show separate lines for the three context conditions: AP (black thin line), PA (blue dashed line), and P (pink thick line).

Proportion of looks. Figures 4 and 5 show the average proportion of looks to agent (leftmost panels), patient (center panels), and instrument (right panels) for all three context conditions for children (top row) and adults (bottom row). With a congruent picture on the screen (Figure 4), adults look to the agent before the pronoun onset, but move their gaze to the patient after hearing the pronoun. Children's gazes follow roughly the same pattern, but show different patterns for the different context conditions. A clear example of children's and adults' different reactions on context is found in the looks to the correct antecedent (patient) in Figure 5: Upon hearing the pronoun, adults move their eyes away from the agent towards the correct antecedent. This pattern is strongest for the AP context, and least strong for the single-referent (P) context. Children show only a small increase in looks to the correct antecedent and least of all for the AP context condition.

GAMM analysis. A logistic GAMM was used for analysis of the gaze data. Logistic regression methods allow as input a two column matrix with the number of looks to the target area of interest in one column (i.e., the number of successes) and the number of looks to the other areas of interest (i.e. the number of failures) in the other column. We prefer to analyze the count data with a logistic (nonlinear) regression analysis over analyzing proportions of looks or logits, because these measures do not take into account that the variance in binomial data is not independent of the mean.[1]

All interactions between the context (AP, PA, or P), area of interest (agent, patient, instrument), image type (congruent or incongruent with a pronoun sentence) and age group (Adults, Children) were combined in one predictor. Each contrast was included as a nonlinear interaction with *Time*, to test whether the different conditions showed variations in gaze pattern over time. In addition, nonlinear random effects over *Time* for participants and items were included, and a random intercept for each individual trial (unique Subject-Item combination). Because a logistic GAMM of gaze data takes a very long time to run, we did not use a model comparison procedure to prune down the model. Instead, the predictions generated from the full model were used to test the differences that were not coded in the contrasts directly. The best fitting model is presented in Appendix 1.

Context effects

The different context conditions result in significantly different gaze patterns for adults and children. Figure 6 focuses on the differences in the 1500 ms window after pronoun onset. It shows the differences in children's and adults' gaze patterns between the AP, PA, and

[1] We used counts per 200 ms timebins as dependent variable rather than a binary value (i.e., looking to area of interest or not) for each measurement, because *a*) counts reduce the data size without loosing much information; and *b*) binning the data reduces the correlation between subsequent measurements, which in turn may reduce the autocorrelation in the residuals (van Rij et al., submitted). The size of the timebins was set on 200 ms, because it takes around 200 ms to plan a saccade (e.g., Rayner, Slowiaczek, Clifton, & Bertera, 1983).

P conditions as estimated by the statistical model. The horizontal lines show where the context conditions differ significantly (i.e., 95% confidence interval of the difference curve does not include zero). Thus, both age groups are sensitive to the structure of the preceding context.

Figure 6 also suggests a difference in timing of looks towards the correct antecedent between the AP and PA context conditions. The antecedent is more prominent in the PA context, because it is mentioned first. This may cause participants to quickly move their gaze towards the correct antecedent when a congruent picture is displayed. In contrast, in the AP context the agent is more prominent, which may result in more competition and longer processing time. With an incongruent picture, in which the agent is highly prominent, both conditions peak later than with a congruent picture, but the PA context even later than AP.

Interestingly, for children the differences between the two-referent contexts AP and PA are estimated larger than the differences between the AP and P contexts. For adults, on the other hand, the difference between the AP and PA contexts is smaller than the difference between the AP and P contexts, because the AP context generally generates more looks to the correct antecedent (which is the patient) than the PA context, and the

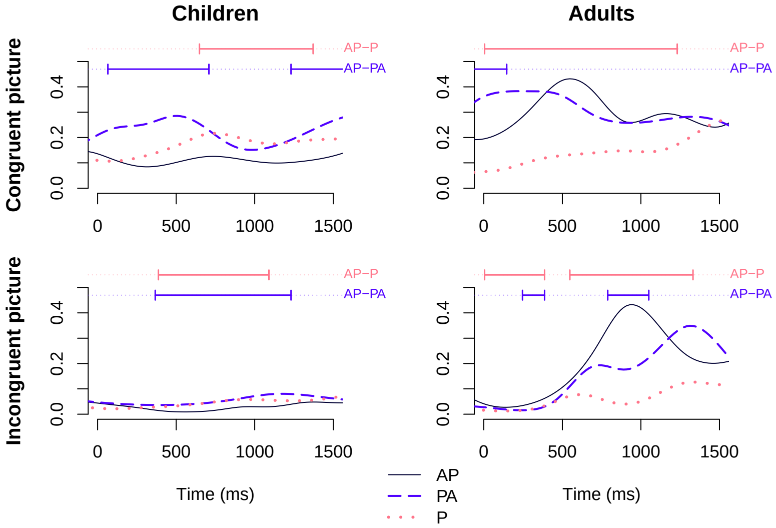


Figure 6. Comparison of context conditions. Estimated proportion of looks to the correct antecedent in contexts AP (thin solid black line), PA (dashed blue line), and P (thick dotted red line). The horizontal intervals at the top of each plot indicate significant differences (95%CI of difference does not include zero) between the AP and P context (highest horizontal interval) and between the AP and PA context (lower horizontal interval).
The estimated effects (model predictions) are transformed back to proportions after removing random effects of participants and items.

PA context generally generates more looks to the correct antecedent than the P context (as is also visible in Figures 4 and 5). This suggests that children react differently to the structure of the context than adults.

Comparison adults and children

A direct comparison between adults' and children's gaze pattern is presented in Figure 7 and Figure 8. These estimates are also derived from the logistic GAMM model that fitted the gaze patterns, but now the patterns of adults and children are compared, rather than the different context conditions. Figure 7 shows the comparisons for the congruent picture (other-oriented action). When we focus on looks to the correct antecedent (patient, bottom row), it is clear that adults look more to this referent than children in the AP condition (significant between 300 and 600 ms after pronoun onset), but in the PA and the P context conditions, there is no difference between children and adults after pronoun onset. Only after the end of the sentence (dashed vertical line), children start to look more to the correct antecedent in the PA context condition.

Figure 8 shows the comparisons for the incongruent picture (self-oriented action). Adults' increased looks to the patient, as shown in Figure 6, is significantly different from the children's gazes after an AP context (between 400 and 1500 ms), and after a PA context

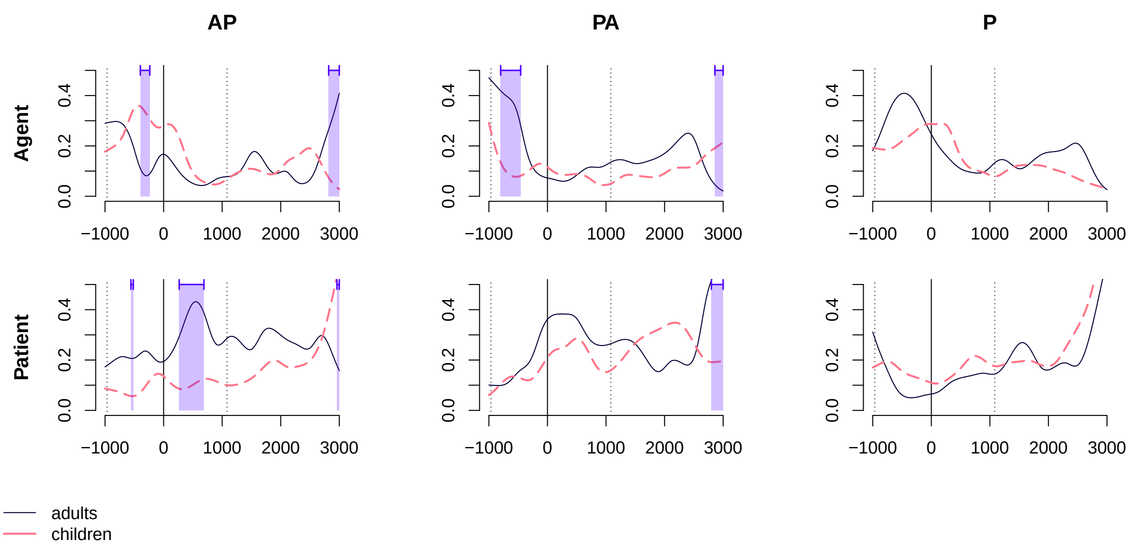


Figure 7. Children versus adults. Comparing the gaze data of children (pink dashed line) and adults (black solid line) for **congruent pictures** with an other-oriented action. The blue areas indicate significant differences (95%CI of difference does not include zero). The dashed vertical lines indicate the average onset and offset of the sentence with the pronoun. The estimated effects (model predictions) are transformed back to proportions after removing random effects of participants and items.

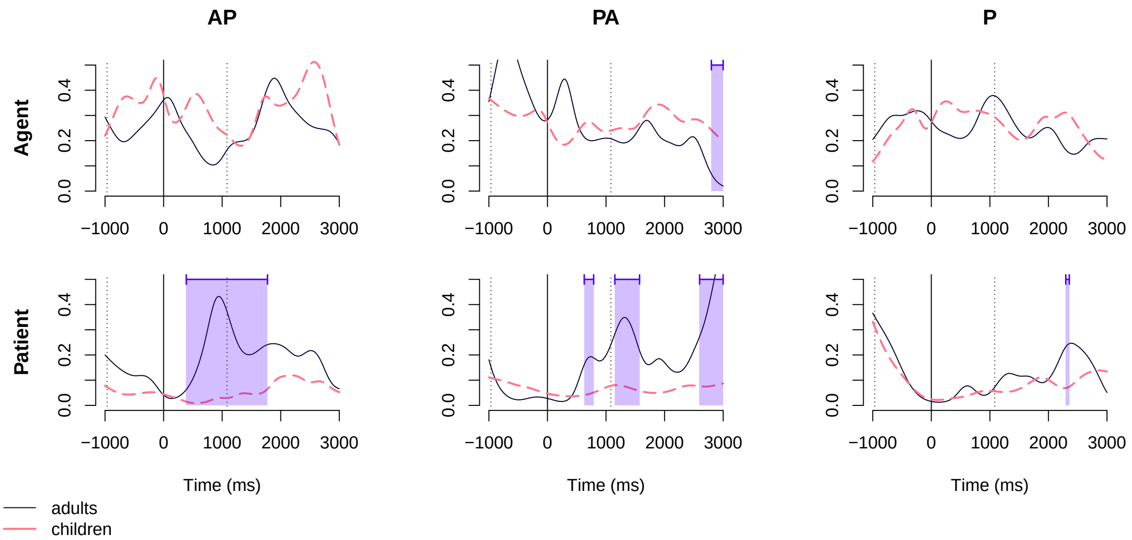


Figure 8. Children versus adults. Comparing the gaze data of children (pink dashed line) and adults (black solid line) for **incongruent pictures** with a self-oriented action. The blue areas indicate significant differences (95%CI of difference does not include zero). The dashed vertical lines indicate the average onset and offset of the pronoun sentence. The estimated effects (model predictions) are transformed back to proportions after removing random effects of participants and items.

(between 1200 and 1500 ms). Late differences (later than 2000 ms after pronoun onset) are found for the PA and P conditions.

Thus, the gaze data shows that children are sensitive to the structure of the discourse context, because their gaze varies with context. Most differences between children and adults just after hearing the pronoun are found in the AP context condition, and children do not differ much from adults in where they look at the screen in the P context condition.

Discussion

This study investigated how the structure of the discourse context influences children's and adults' processing of object pronouns. The behavioral results replicated the finding that children allow a co-referential interpretation for pronouns, whereas for adult participants the pronoun could only refer to the referent that was not the local subject. Children's interpretation of reflexives was more adult-like, indicated by a higher sensitivity and no bias. Children's sensitivity for pronouns increased when only the antecedent of the pronoun was introduced in the preceding context sentence (cf. Spenader et al., 2009). Thus, increasing the prominence of the antecedent guides children in the right direction.

The gaze data shows that children and adults are influenced by the preceding context during pronoun processing, because they showed a different pattern in each context condition. Comparing children's and adults' gaze patterns directly, we found that children show a similar gaze pattern as adults after a single-referent context (P). This aligns with the behavioral results in which children showed more adult-like performance on pronoun interpretation with the single-referent context than with the two-referent contexts. Children and adults react differently on an AP context. Adults look significantly more to the antecedent after an AP context in comparison with the other contexts. Children, on the other hand, look significantly less to the antecedent after an AP context in comparison with the other contexts. In the PA context condition, children and adults do not show significant differences in their gaze during sentence processing, but only after the sentence is finished. Together, the behavioral responses and the gaze data suggest that children use context information to resolve the pronoun.

Discourse prominence

Children show adult-like performance after hearing a P context, but show different gaze patterns than adults after an AP context. In the AP context, the agent is more prominent than the antecedent of the pronoun, because it is introduced first and mentioned again in subject position (e.g. Gernsbacher & Hargreaves, 1988; Järvikivi et al., 2005). Given that children allow a pronoun to have a co-referential interpretation (e.g., Chien & Wexler, 1990; Koster, 1993; Spenader et al., 2009), we argue that children do not look to the correct antecedent in the AP context condition because they take the most prominent referent to be the antecedent of the pronoun, which is the agent.

The difference between the two-referent contexts, with more adult-like gaze patterns after a PA context than after an AP context, suggests that discourse prominence rather than discourse coherence influences children's gazes. In the AP context, the referent that was introduced in first position continued as topic of the pronoun sentence, but in the PA context the topic shifts to the referent that was not the firstly introduced referent, but rather was introduced later. Thus, the AP context provides a more coherent discourse than the PA context, and hence should be easier to process from a discourse coherence perspective. However, children show more adult-like performance after a PA context than after an AP context.

Not only the discourse context seems to affect the referent's prominence, but also the visual context. With an incongruent picture showing a self-oriented action, children's looks to the antecedent are reduced in all conditions, but most strongly in the AP context condition. In a self-oriented action the agent is visually more prominent than the patient, who is merely watching the agent perform the action.

However, some effects are not easy to explain on the basis of discourse prominence: (*1*) Children and adults look more to the antecedent in the PA context condition than in the P context condition, although the antecedent is more prominent in the P context condition; (*2*) With an incongruent picture, adults look more to the antecedent when the antecedent is less prominent. This may be due to the task, in which participants are asked to judge whether the sentence is a correct description of the picture. Lower prominence of the patient may lead to increased looks to the patient to be sure to reject the interpretation shown on the screen. Thus, the gaze pattern suggests a more complex interaction between grammar, discourse prominence, and visual context.

Eye tracking

The gaze results shed more light on the study of Bergmann et al. (2011), who showed that 4-year-old children looked more to the antecedent after hearing a sentence with a pronoun than after hearing a sentence that made no reference to any of the depicted figures. They conclude that children already possess the linguistic knowledge required for object pronoun interpretation from 4 years old on, but that task difficulties are responsible for the observed delayed comprehension of object pronouns in comparison with reflexives (cf. Conroy et al., 2009). Our data, however, suggests that the amount of looks to the antecedent relative to the looks to the agent is not necessarily informative of the linguistic knowledge being used, because the amount of looks to the agent and patient is influenced by the discourse context and visual context. In contrast to the children in the study of Bergmann et al. (2011), the adults in our experiment looked equally often or significantly less to the patient than to the agent in some conditions (this is most clear in the single-referent context). However, they did not show ambiguity of object pronouns in their responses. We argue that not the amount of looks to a certain area of interest, but rather the comparison with adults' gaze data should be used for drawing conclusions about children's linguistic knowledge.

Our study supports the conclusion of earlier studies that gaze pattern may reveal adult-like linguistic performance while the behavioral data does not show adult-like performance (cf. Bergmann et al., 2011; Sekerina et al., 2004): in the PA context condition, children's gaze pattern does not differ significantly from adults' gaze data in the 2000 ms after pronoun onset. Children's behavioral responses on that condition, however, show a much lower sensitivity and a larger response bias than adults' behavioral response data and do not differ significantly from their behavioral responses in the AP context condition. Our results are also in line with Clackson et al. (2011), who report that children and adults are distracted by ungrammatical gender-matching referents during object pronoun processing. Although looks to the agent generally decrease after the onset of the pronoun, a considerable amount of agent looks are measured early after the pronoun.

Interestingly, children's reactions to the pronoun in their gaze behavior are not much slower than adults' gaze behavior, similar to other eye-tracking studies of object pronoun processing (Bergmann et al., 2011; Clackson et al., 2011). This contrasts with the results of Sekerina et al. (2004), who found that children's gaze behavior was around 1000 ms slower than adults' gaze behavior on the interpretation of ambiguous object pronouns in prepositional phrases. This difference might be caused by the slightly different constructions that were being investigated, but may also be due to the different task. Sekerina et al. (2004) presented two pictures on the screen and children had to point to the picture that illustrated the correct interpretation. In our task, only one picture was being presented. The relatively early effects fit in with earlier findings on incremental sentence processing (e.g., Altmann & Kamide, 1999).

Conclusion

The current study shows that discourse information can direct children towards an adult-like interpretation of object pronouns or towards an incorrect co-referential interpretation. Increasing the saliency of the antecedent in the discourse context resulted in more adult-like performance in children's responses and gaze patterns. However, when the incorrect co-referential referent was salient in the visual context and the linguistic context, children ignore the correct antecedent, whereas adults show increased looks to the antecedent. This confirms that children's interpretation of pronouns is not yet fully constrained by the grammar, but is partially based on other sources of information, such as the discourse context and visual context.

References

Boehm, U., van Maanen, L., Forstmann, B., & van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *Neuroimage*, *96*(95–105).

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. C. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527-565.

Başkent, D., van Rij, J., Ng, Z. Y., Free, R., & Hendriks, P. (2013). Perception of spectrally degraded reflexives and pronouns by children. *The Journal of the Acoustical Society of America*, *134*(5), 3844–3852.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748-769.

Bergmann, C., Paulus, M., & Fikkert, P. (2011). Preschoolers' comprehension of pronouns and reflexives: the impact of the task. *Journal of Child Language*, *39*(4), 777–803.

Chien, Y. C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition*, *1*(3), 225-295.

Chomsky, N. (1981). *Lectures on government and binding: the Pisa lectures.* Dordrecht, The Netherlands: Foris Publications.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in english: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*, 128-144.

Conroy, A., Takahashi, E., Lidz, J., & Phillips, C. (2009). Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry*, *40*(3), 446–486.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

De Villiers, J., Cahillane, J., & Altreuter, E. (2006). What can production reveal about Principle B? In K. Deen, J. Nomura, B. Schulz, & B. Schwartz (Eds.), *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition-North America* (Vol. 1, p. 89-100). Honolulu, HI.: University of Connecticut Occasional Papers in Linguistics 4. (The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition-North America, University of Connecticut Occasional Papers in Linguistics)

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*(6), 699-717.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311-347.

Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry*, *24*, 69–102.

Hendriks, P. (2014). *Asymmetries between language production and comprehension.* Springer, Dordrecht.

Hendriks, P., Banga, A., van Rij, J., Cannizzaro, G., & Hoeks, J. C. J. (2011). Adults' on-line comprehension of object pronouns in discourse. In A. Grimm, A. Müller, C. Hamann, & E. Ruigendijk (Eds.), *Production-comprehension asymmetries in child language.* (pp. 193–216). De Gruyter Mouton, Berlin.

Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience*, *29*(7), 877–892.

Järvikivi, J., van Gompel, R. P., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16*(4), 260–264.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*(5), 709-748.

Koster, C. (1993). *Errors in anaphora acquisition* (Unpublished doctoral dissertation).

Lin, X., & Zhang, D. (1999). Inference in generalized additive mixed modelsby using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*, 381–400.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide.* Psychology Press.

Masson, M. E. J., & Kliegl, R. (2013). Modulation of additive and interactive effects on lexical decision by trial history. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 898–914.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2009). Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics*, *20*(3), 599-626.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5-19.

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (submitted). Eye movements reflect acoustic cue informativity and statistical noise.

Philip, W., & Coopmans, P. (1996). The double Dutch delay of Principle B effect. In *Annual boston university conference on language development* (Vol. 20).

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar.* Malden, MA: Blackwell.

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*, *25*(1), 115–129.

R Core Team. (2014). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 912–922.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition*, *89*(1), B1–B13.

Sekerina, I. A., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language*(123–152).

Song, H.-J., & Fisher, C. (2005). Who's "she"? discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language*, *52*(1), 29-57.

Song, H.-J., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua*, *117*(11), 1959-1987.

Spenader, J., Smits, E. J., & Hendriks, P. (2009). Coherent discourse solves the Pronoun Interpretation Problem. *Journal of Child Language*, *36*(1), 23-52.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137–149.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Intergration of visual and linguistics information in spoken language comprehension. *Science*, *268*, 1632–1634.

Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis, and interpretation in child grammar.* MIT Press.

Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). Frequency effects on the articulation of german i and u: evidence from articulography. In *Interspeech* (pp. 1302–1306).

van Rij, J., Natalya, P., van Rijn, H., Wood, S. N., & Baayen, R. H. (submitted). Pupil dilation to study cognitive processing: challenges and solutions for time course analyses.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs [Computer software manual]. Retrieved from http://cran.r-project.org/package=itsadug (R package version 0.9)

van Rij, J., van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language*, *37*(3), 731-766. (Chapter 2 of this thesis)

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Appendix 1: Statistical models

The data, analyses, and R scripts are reported in the Supplementary Materials, which are available on `www.sfs.uni-tuebingen.de/~jvanrij/paperdata/contextgazedata.html`(Temporary address, will be replaced upon acceptance.). The Supplementary Materials also provides more information on the Signal Detection Theory (e.g. Macmillan & Creelman, 2004) measures used in this paper.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	1.4562	0.0814	17.8888	$<$ 0.0001
expressionP	-0.4190	0.0781	-5.3646	$<$ 0.0001
contextAP	-0.1573	0.0730	-2.1555	0.0316
contextPA	-0.1107	0.0730	-1.5167	0.1300
measureC	-1.5437	0.0813	-18.9935	$<$ 0.0001
expressionP:contextAP	-0.0691	0.0844	-0.8190	0.4132
expressionP:contextPA	-0.2455	0.0844	-2.9091	0.0038
expressionP:measureC	0.2416	0.0690	3.4988	0.0005
contextAP:measureC	0.2061	0.0843	2.4446	0.0148
contextPA:measureC	0.2078	0.0843	2.4649	0.0140
B. smooth terms	edf	Ref.df	F-value	p-value
s(AgeMonths):exprAP,contextP,C	1.0000	1.0000	0.0072	0.9324
s(AgeMonths):exprP,contextP,C	1.0000	1.0000	0.0077	0.9301
s(AgeMonths):exprPA,contextP,C	1.0000	1.0000	0.2087	0.6479
s(AgeMonths):exprAP,contextR,C	1.0000	1.0000	0.0472	0.8281
s(AgeMonths):exprP,contextR,C	1.0000	1.0000	0.8890	0.3462
s(AgeMonths):exprPA,contextR,C	1.0000	1.0000	0.0051	0.9430
s(AgeMonths):exprAP,contextP,dp	8.0107	8.6963	3.3023	0.0008
s(AgeMonths):exprP,contextP,dp	3.0854	3.6878	2.1510	0.0790
s(AgeMonths):exprPA,contextP,dp	6.9027	7.9300	1.3419	0.2204
s(AgeMonths):exprAP,contextR,dp	2.3216	2.8670	6.8762	0.0002
s(AgeMonths):exprP,contextR,dp	1.9631	2.3568	4.3461	0.0100
s(AgeMonths):exprPA,contextR,dp	1.7553	2.1392	7.5290	0.0005
s(Subject) [random intercept]	19.4855	38.0000	17.9902	0.0001
s(measure,Subject) [random slope]	34.8061	76.0000	3.7802	0.0019
s(expression,Subject) [random slope]	31.8983	76.0000	2.9258	0.0019

Table 1

Summary of the best-fitting model of behavioral responses, resulting from the model-comparison procedure. The statistics listed in the summary of smooth terms (B.) indicates whether the smooth is significantly different from 0. Factor is a 12-level predictor that captures the interaction between Context (AP, PA, P), Referring Expression (R, P) and Measure (dp, C).

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-2.0286	0.5026	-4.0367	0.0001
FactorP.Adults.AgentFix.OOim	0.4479	0.3037	1.4748	0.1403
FactorPA.Adults.AgentFix.OOim	0.2420	0.3457	0.7000	0.4839
FactorAP.Children.AgentFix.OOim	0.2008	0.5976	0.3360	0.7368
FactorP.Children.AgentFix.OOim	0.1607	0.5734	0.2803	0.7793
FactorPA.Children.AgentFix.OOim	-0.2063	0.6019	-0.3427	0.7318
FactorAP.Adults.PatientFix.OOim	1.0060	0.7115	1.4138	0.1574
FactorP.Adults.PatientFix.OOim	0.3132	0.6851	0.4572	0.6475
FactorPA.Adults.PatientFix.OOim	0.9146	0.7099	1.2883	0.1976
FactorAP.Children.PatientFix.OOim	0.1446	0.6994	0.2068	0.8362
FactorP.Children.PatientFix.OOim	0.6395	0.6732	0.9499	0.3422
FactorPA.Children.PatientFix.OOim	0.6966	0.7015	0.9931	0.3207
FactorAP.Adults.AgentFix.SOim	0.9111	0.4012	2.2712	0.0231
FactorP.Adults.AgentFix.SOim	0.9347	0.3444	2.7140	0.0066
FactorPA.Adults.AgentFix.SOim	0.9792	0.3941	2.4847	0.0130
FactorAP.Children.AgentFix.SOim	1.3039	0.6276	2.0775	0.0378
FactorP.Children.AgentFix.SOim	0.9791	0.5908	1.6572	0.0975
FactorPA.Children.AgentFix.SOim	1.0478	0.6236	1.6802	0.0929
FactorAP.Adults.PatientFix.SOim	0.3419	0.7153	0.4781	0.6326
FactorP.Adults.PatientFix.SOim	-0.2813	0.6894	-0.4080	0.6833
FactorPA.Adults.PatientFix.SOim	-0.0218	0.7113	-0.0306	0.9756
FactorAP.Children.PatientFix.SOim	-1.0095	0.7037	-1.4346	0.1514
FactorP.Children.PatientFix.SOim	-0.5308	0.6735	-0.7881	0.4306
FactorPA.Children.PatientFix.SOim	-0.6775	0.6997	-0.9682	0.3329
B. smooth terms	edf	Ref.df	F-value	p-value
s(Timebin):FactorAP.Adults.AgentFix.OOim	16.3846	17.6068	206.4741	< 0.0001
s(Timebin):FactorP.Adults.AgentFix.OOim	14.7361	16.5684	129.5459	< 0.0001
s(Timebin):FactorPA.Adults.AgentFix.OOim	14.3341	16.3688	120.6453	< 0.0001
s(Timebin):FactorAP.Children.AgentFix.OOim	14.8383	16.7327	148.5872	< 0.0001
s(Timebin):FactorP.Children.AgentFix.OOim	13.8471	15.9239	62.1072	< 0.0001
s(Timebin):FactorPA.Children.AgentFix.OOim	13.9785	16.1815	110.2147	< 0.0001
s(Timebin):FactorAP.Adults.PatientFix.OOim	14.7990	16.6674	70.7508	< 0.0001
s(Timebin):FactorP.Adults.PatientFix.OOim	15.1044	16.8370	132.4441	< 0.0001
s(Timebin):FactorPA.Adults.PatientFix.OOim	14.8384	16.7391	77.9885	< 0.0001
s(Timebin):FactorAP.Children.PatientFix.OOim	14.1087	16.2255	116.2094	< 0.0001
s(Timebin):FactorP.Children.PatientFix.OOim	13.0822	15.3503	65.0810	< 0.0001
s(Timebin):FactorPA.Children.PatientFix.OOim	13.5785	15.8581	81.3340	< 0.0001
s(Timebin):FactorAP.Adults.AgentFix.SOim	14.3610	16.4819	84.0891	< 0.0001
s(Timebin):FactorP.Adults.AgentFix.SOim	14.3094	16.3302	70.9298	< 0.0001
s(Timebin):FactorPA.Adults.AgentFix.SOim	15.9094	17.3970	216.9964	< 0.0001
s(Timebin):FactorAP.Children.AgentFix.SOim	15.9265	17.6060	136.5297	< 0.0001
s(Timebin):FactorP.Children.AgentFix.SOim	15.7525	17.4014	119.0902	< 0.0001
s(Timebin):FactorPA.Children.AgentFix.SOim	13.1662	15.4310	56.8842	< 0.0001
s(Timebin):FactorAP.Adults.PatientFix.SOim	15.4606	17.2598	171.9854	< 0.0001
s(Timebin):FactorP.Adults.PatientFix.SOim	16.7818	18.0228	307.3906	< 0.0001
s(Timebin):FactorPA.Adults.PatientFix.SOim	16.1474	17.6687	262.9189	< 0.0001
s(Timebin):FactorAP.Children.PatientFix.SOim	14.8541	16.9355	95.5626	< 0.0001
s(Timebin):FactorP.Children.PatientFix.SOim	14.2157	16.3815	118.9343	< 0.0001
s(Timebin):FactorPA.Children.PatientFix.SOim	9.4205	11.6024	24.5838	0.0142
s(Timebin,fixSubj)	1281.1974	1382.0000	5692209.5737	< 0.0001
s(Timebin,fixItem)	397.5968	430.0000	2322134.9160	< 0.0001
s(fixRec)	1841.8878	2092.0000	71804.7968	< 0.0001

Table 2

Summary of the full model of gaze data on sentences with a pronoun. Fix consists of a combination (cbind) of success counts and failure counts per time bin. Factor is a 24-level predictor that captures the interaction between Context (AP, PA, P), Age group (children, adults), AOI (agent, patient), and image type (other-oriented, self-oriented).