Journal of Experimental Child Psychology 173 (2018) 250–267

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Individual differences in children's pronoun processing during reading: Detection of incongruence is associated with higher reading fluency and more regressions

Sarah Eilers *, Simon P. Tiffin-Richards, Sascha Schroeder

Max Planck Research Group REaD, Max Planck Institute for Human Development, 14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 26 July 2017
Revised 27 March 2018

Keywords:
Pronoun resolution
Sentence processing
Children's reading
Reading development
Eye tracking
Comprehension monitoring

ABSTRACT

In two eye tracking experiments, we tested fourth graders' and adults' sensitivity to gender feature mismatches during reading of pronouns and their susceptibility to interference of feature-matching entities in the sentence. In Experiment 1, we showed children and adults two-phrase sentences such as "Leon{m}/Lisa {f} shooed away the sparrow{m}/the seagull{f} and then he{m} ate the tasty sandwich." Eye tracking measures showed no qualitative differences between children's and adults' processing of the pronouns. Both age groups showed longer gaze durations on subject mismatching than on matching pronouns, and there was no evidence of interference of a gender-matching object. Strikingly, in contrast to the adults, not all fourth graders reported detection of the subject gender mismatch. In Experiment 2, we replicated earlier results with a larger sample of children ($N = 75$) and found that only half of the fourth graders detected the gender mismatch during reading. The detectors' reading pattern at the pronoun differed from that of the non-detectors. Children who reported detection of the mismatch showed a reading pattern more similar to the adults. Children who did not report detection of the mismatch had comparably slower gaze durations and were less likely to make regressions directly at the pronoun. We conclude that children

* Corresponding author.
E-mail address: eilers@posteo.de (S. Eilers).

https://doi.org/10.1016/j.jecp.2018.04.005

who read more fluently use their available processing resources to immediately repair grammatical inconsistencies encountered in a text.

Introduction

Reading is a complex task that involves not only word decoding but also linking pieces of information across longer text passages. Proficient readers use coherence markers to integrate new information into the current situation model (Zwaan & Radvansky, 1998). Pronouns and other anaphora are important markers for coherence because they link entities across sentences and, thus, serve as a cue for the way in which phrases are interconnected (Ariel, 2004). Online processing and integration of pronouns, therefore, is important for situation model building and ultimately for text comprehension (Garnham, Oakhill, & Johnson-Laird, 1982). However, pronouns are semantically underspecified because they carry only number and gender information. Therefore, they often have to be inferred based on the sentence context (Kehler, 2002; Kehler, Kertz, Rohde, & Elman, 2008).

In the current study, we investigated whether the online processing of pronominal gender information is a possible source of reading difficulty for children. In Experiment 1, we tested adults' and children's sensitivity to gender feature mismatches on the pronoun and their susceptibility to interference effects when a gender-matching object is present. In Experiment 2, we compared the online pronoun processing of children who reported detection of the gender mismatch with that of children who did not report detection. We were interested in the eye movement patterns associated with the report of mismatch detection and inter-individual differences that may contribute to successful mismatch detection in children.

Online pronoun resolution in proficient readers

Proficient readers infer the antecedent of a pronoun online by combining lexical information (e.g., gender of the pronoun) and contextual information (e.g., verb meaning, disambiguating sentence information). In a self-paced reading experiment, Garnham and Oakhill (1985) showed that adults need more time for the integration of a subclause when there is no gender cue on the pronoun and the antecedent needs to be inferred entirely from context. This shows that readers use the gender cue on the pronoun for resolution online during reading. Gender mismatches of pronoun and antecedent, therefore, should disrupt the reading process. In a self-paced reading study with proficient readers, Carreiras, Garnham, Oakhill, and Cain (1996) found longer reading times for the last sentence of a story when it contained a mismatching pronoun for the stereotypical gender of a referent (i.e., female for nurse, male for doctor). They concluded that adults use gender information as soon as it becomes available, and their results show that adult readers form expectations for the gender of a pronoun, such that gender mismatches result in longer processing times. In a study with a stronger manipulation, Rigalleau, Caplan, and Baudiffier (2004) presented adults with sentences such as "Wendy complimented Nancy because she/*he made an effort" and found that reading times on the subordinate clause were significantly longer when the pronoun did not match the two antecedents. Their results also show that the gender feature is a strong determinant for the identification of an antecedent. When there is no available gender-matching antecedent, processing of the pronoun is made difficult to a point where a proficient reader does not engage in resolution at all even if enough context information is available to infer the correct antecedent. In the experiment by Rigalleau et al., however, response accuracy on the comprehension questions of the gender-mismatch sentences was equal to that of the gender-match sentences. However, response latencies were significantly longer for the mismatch sentences. This suggests that although readers understood the sentences despite the mismatching pronoun, a gender-mismatching pronoun has both immediate and substantial effects on the efficiency of sentence comprehension.

Children's comprehension and processing of anaphora

Pronoun processing and comprehension have been studied extensively in very young children using listening tasks. Although this line of research has produced mixed results due to a variety of methods, materials, age groups, and languages studied (Hickmann, Schimke, & Colonna, 2015), many studies provide evidence that children can use gender cues effectively to establish pronoun–antecedent relationships during listening from 3 years of age (e.g., Arnold, Brown-Schmidt, & Trueswell, 2007; see also Sekerina, 2015, for a review). Therefore, we have reason to assume that by the time children reach the end of primary school, they are able to identify the correct referent for a pronoun online in spoken language. However, this is not the case for reading. It has repeatedly been shown that children struggle to name the correct referent for pronouns when reading a text. Yuill and Oakhill (1988) tested 7- and 8-year-olds' comprehension of sentences containing a pronoun and one or two gender-matching antecedents. Depending on the difficulty of the pronominal inference, children had an error rate of up to 28%. In a similar experiment, Oakhill and Yuill (1986) showed that 7- and 8-year-olds have difficulties in naming the correct referent for the personal pronoun after having read sentences such as "Peter lent ten pence to Tom [Liz] because he [she] was very poor." The children performed significantly worse when there was no gender cue (16–27% error rate) compared with when there was a gender cue (2–14% error rate). These studies show that pronoun resolution is difficult for children and that they do rely on gender cues for comprehension. However, studies targeting comprehension cannot clarify whether children use the cue spontaneously online or whether they use it for offline comprehension only when prompted by a question.

It is conceivable that children's online processing of pronouns remains "shallow" because reading is more effortful for children in general, given that their word decoding is less automatized than that of proficient readers (e.g., Gagl, Hawelka, & Wimmer, 2015). Contrasting the reading process in adults and children using eye tracking, it has been shown that children make more and longer fixations than adults and engage in more unselective rereading (for reviews, see Blythe & Joseph, 2011; Schroeder, Hyönä, & Liversedge, 2015). In studies of inter-individual differences in children's reading, slow decoding has been associated with poor comprehension (Nation, 2005; Nation & Snowling, 1998). Children may struggle in particular when they need to integrate information that spans longer distances of text online. In a seminal eye tracking experiment with 10-year-olds, Joseph, Bremner, Liversedge, and Nation (2015) compared the processing of anaphora with typical antecedents (a truck–the vehicle) and atypical antecedents (a crane–the vehicle) in short texts where the anaphor (the vehicle) was either near or far from its antecedent. The authors showed that when resolution is most difficult (i.e., atypical anaphors that are far from the antecedent), children might not resolve the anaphor during reading at all. In other words, in the difficult condition, children might not understand that "the crane" and "the vehicle" refer to the same entity in their situation model, resulting in impoverished comprehension. Studies on the resolution of pronouns in children are still rare. Recent results from a study contrasting pronouns with repeated names indicate that 8-year-olds already show a repeated name penalty effect during online reading. This suggests that beginning readers show the same discourse-level expectations for pronouns as adults (Eilers, Tiffin-Richards, & Schroeder, 2018). However, it is still unclear how children resolve the pronoun, specifically which type of information they take into account and whether resolution happens online in children's reading.

Because there are large inter-individual differences in children's reading comprehension and related component skills (e.g., Cain, Oakhill, & Bryant, 2004; Oakhill, 1982, 1984; see also Nation, 2005, for a review), children may differ in their ability to resolve pronouns online. A recent reading comprehension study with children indeed demonstrated that the ability to resolve pronouns correctly may account for individual variance in reading comprehension (Elbro, Oakhill, Megherbi, & Seigneuric, 2017).

As one of few studies targeting online processing of pronouns in children, Ehrlich, Rémond, and Tardieu (1999) investigated 10-year-olds' resolution of pronouns in expository text passages. They conducted a self-paced reading experiment in which the children could decide to reread previous parts of the text via button press. The authors demonstrated that less skilled comprehenders—that is, children who read as fluently as their peers but perform poorly in a standard reading comprehension test—struggled when encountering personal pronouns as opposed to repeated names in expository texts. Reading times were longer for less skilled comprehenders than for skilled comprehenders overall; however, reading times on sentences containing a pronoun were longer for skilled comprehenders but not for less skilled comprehenders. The skilled comprehenders further chose to reread earlier parts of the text more often than the less skilled comprehenders. Ehrlich et al.'s study demonstrates that skilled readers among the children may display reading patterns that are qualitatively different from those of less skilled readers. However, because the study did not address the time course of the resolution process, it is unclear where the specific problems of less skilled comprehenders arise during pronoun processing.

In an early eye tracking study with 8-year-old children, Murray and Kennedy (1988) further showed that the eye movement behavior associated with pronoun resolution in skilled comprehenders differs from that in less skilled comprehenders. Skilled comprehenders made regressions more selectively when reading sentences containing pronouns, whereas less skilled comprehenders made shorter, less selective regressions (termed "backtracking" by the authors). The study suggests that selective regressions in children may be associated with better comprehension. In addition, if children use gender information for the association of pronoun and antecedent online, they may be distracted by an interfering gender-matching antecedent. In German, all nouns carry linguistic gender (for an overview, see Fagan, 2009). Linguistic gender is indicated by its preceding article, *der* for male nouns (e.g., *der Spatz* [the sparrow], *der Brief* [the letter]) and *die* for female nouns (e.g., *die Möwe* [the seagull], *die Karte* [the card]). Reference to a mouse in German requires the personal pronoun *sie*, whereas reference to a hamster requires the pronoun *er*. The gender of a noun is internalized early during language acquisition in German. A number of studies have shown that pronoun processing may be inhibited by interfering gender–feature matching discourse entities in proficient readers even when these interfering entities are excluded as the antecedent of the pronoun for structural reasons (see Jäger, Engelmann, & Vasishth, 2017, for a review). This shows that the gender feature is a strong determinant for the association of pronoun and antecedent in proficient readers. Children have been shown to resort to shallow processing when confronted with reading material that is difficult for them (e.g., Joseph et al., 2015). For beginning readers, sentences with a mismatching pronoun may be particularly misleading when there is a gender-matching distractor present even if this distractor does not match the pronoun based on sentence context. In other words, children's reading may be sufficiently shallow to allow processing of a mismatching pronoun when the distractor is a gender match for the pronoun.

In summary, our interests in the current study were (a) whether children process mismatching pronouns similarly to adults and the efficiency of their processing, (b) whether children are susceptible to interference effects by a gender-matching distractor, and (c) differences in pronoun processing in children who reported detection of the mismatch compared with children who did not report detection of the mismatch.

Experiment 1

Rationale

We studied children's and adults' processing of pronouns in a mismatch paradigm with gender-mismatching and gender-matching pronouns and an intervening discourse entity of the same or different gender. Our main interests were effects in the pronoun region and the sentence-final region, that is, the region directly following the pronoun. We used gaze duration in the pronoun region as an early indicator of processing difficulty, and regression probability and regression path duration as late indicators of processing difficulty, associated with "repair" processes. We also report total reading time, a measure that incorporates gaze duration and rereading time. In the final region, we were interested in the integration processes typically found at the end of a sentence. Effects in this region may be related to pronoun resolution if resolution processes affect regions downstream from the anaphor. More important, however, readers are expected to engage in sentence-final meaning integration.

This process is informative in mismatch paradigms because it allows conclusions about whether readers repair local inconsistencies online or whether such inconsistencies disrupt their ability to integrate sentence meaning.

We presented two-phrase sentences, such as "Leon/Lisa shooed away the sparrow/the seagull and then he ate the tasty sandwich," to children and adults. In all stimuli, the subject of the first clause (Leon) was the contextually plausible antecedent for the pronominal subject of the second clause (he). In the example above, *he* must be co-referential with *Leon*. The gender–feature mismatch prohibits bonding of the pronoun with *Lisa* (subject gender mismatch). In the pronoun region, we expected longer gaze durations for gender-mismatching pronouns than for gender-matching pronouns for the adults. We further hypothesized that adults would initiate repair processes at the mismatching pronoun, such that regression probability increases with mismatching pronouns and regression path duration is prolonged. For the children, we also hypothesized longer gaze durations for mismatching pronouns than for matching pronouns, indicating mismatch detection. However, for children the effect may be spatially delayed and occur only in the region following the pronoun when meaning integration is expected. Such a delayed effect would be consistent with earlier findings for children's processing of implausible thematic relations (Joseph et al., 2008). Moreover, we hypothesized that children differ from adults in later processing stages. Integration of a mismatching pronoun requires repair of the inconsistent information, and beginning readers might not be able to engage in online repair due to cognitive resource constraints.

In this study, we were also interested in whether children's processing of a pronoun may be influenced by a second gender-matching discourse entity. We expected that the gender-matching object is a potential source of confusion for pronoun processing during reading. We hypothesized that in children the effect of a mismatching subject may be modulated by an object in the sentence that is a gender match to the pronoun (*the sparrow*{m} vs. *the seagull*{f}). Such confusion of the appropriate antecedents during processing would indicate that children rely heavily on gender cues in pronoun integration, arguably because gender information is available directly at the word level, which makes it easier to integrate than context information.

Method

Participants

In Experiment 1, we recruited 29 children from three schools in Berlin. Of these 29 children, 5 were excluded because of missing data due to technical issues, and data from 1 child were excluded because the child was a late immigrant (i.e., arrived in Germany after 5 years of age). Of the remaining 23 children, 2 were early immigrants (arrived in Germany before age 5). All participating children received reading instruction in German only. Of the participating children, 17 were girls. The children were 9 years old (SD = 15 months) on average, and all attended fourth grade. All children had normal or corrected-to-normal vision. We further recruited 25 young adults (M_{age} = 25.24 years, SD = 3.2) from universities in Berlin via mailing lists. Of these adults, 17 were women. All adults were native speakers of German and reported normal or corrected-to-normal vision.

To verify that the samples did not differ from typically developed readers in their age group, we administered a reading fluency test and a reading comprehension test. All participants completed the SLRT-II standardized test of reading fluency (Moll & Landerl, 2010). Children did not differ from the population mean standard score of 50 in word reading fluency (M = 53.0, SD = 24.4), $t(22) < 1$, $p = .56$; neither did the adults (M = 51.1, SD = 29.9), $t(24) < 1$, $p = .86$. Children further completed the ELFE 1–6 standardized reading comprehension test (Lenhard & Schneider, 2006). Because of technical difficulties during data collection, the sentence comprehension subtest scores and summed z scores could not be calculated. The obtained z scores, however, indicate that the children did not differ significantly from the population mean in the word comprehension subtest ($M = -0.07$, SD = 0.80), $t(22) < 1$, $p = .68$, or text comprehension subtest ($M = -0.15$, SD = 1.0), $t(22) < 1$, $p = .46$.

Materials

Materials consisted of 48 sentences as depicted in Table 1. The sentences were written in four experimental versions, resulting in 192 stimuli. All items were written specifically for primary school

Table 1
Two examples of stimulus materials in the four conditions.

					Region of interest		
					Pronoun	Final	Condition
Leon{m}	verjagte	den Spatz{m}	und dann	aß	[er das]	[leckere Brötchen.]	match–match
Leon	*shooed away*	*the sparrow*	*and then*	*ate*	*he the*	*tasty sandwich*	
Max{m}	schrieb	den Brief{m}	und dann	lief	[er zur]	[nächsten Post.]	
Max	*wrote*	*the letter*	*and then*	*ran*	*he to-the*	*next post office*	
Lisa{f}	verjagte	den Spatz {m}	und dann	aß	[er das]	[leckere Brötchen.]	mismatch–match
Lisa	*shooed away*	*the sparrow*	*and then*	*ate*	*he the*	*tasty sandwich*	
Mia{f}	schrieb	den Brief{m}	und dann	lief	[er zur]	[nächsten Post.]	
Mia	*wrote*	*the letter*	*and then*	*ran*	*he to-the*	*next post office*	
Leon{m}	verjagte	die Möwe{f}	und dann	aß	[er das]	[leckere Brötchen.]	match–mismatch
Leon	*shooed away*	*the seagull*	*and then*	*ate*	*he the*	*tasty sandwich*	
Max{m}	schrieb	die Karte{f}	und dann	lief	[er zur]	[nächsten Post.]	
Max	*wrote*	*the card*	*and then*	*ran*	*he to-the*	*next post office*	
Lisa{f}	verjagte	die Möwe{f}	und dann	aß	[er das]	[leckere Brötchen.]	mismatch–mismatch
Lisa	*shooed away*	*the seagull*	*and then*	*ate*	*he the*	*tasty sandwich*	
Mia{f}	schrieb	die Karte{f}	und dann	lief	[er zur]	[nächsten Post.]	
Mia	*Wrote*	*the card*	*and then*	*ran*	*he to-the*	*next post office*	

Note. Literal English translations are given to ease interpretation. Brackets indicate regions of interest for analysis. {m}, male gender; {f}, female gender.

children and contained concepts to which they could be expected to relate. All items consisted of a complex clause composed of a main clause and a subordinate clause. The main clause always introduced a boy or girl by name in subject position and either an animal or an artifact in object position. The protagonists in subject position could be of either male or female gender (e.g., *Leon* vs. *Lisa, Max* vs. *Mia;* see below). The animal or artifact in object position could be of either male or female linguistic gender (e.g., *der Spatz* vs. *die Möwe, der Brief* vs. *die Karte*).

The sentences consisted of 11 or 12 words, with 2 words in each region of interest (see Table 1). The pronoun region contained a pronoun followed by a preposition or an article to increase the likelihood of a fixation in this region. Note that the subordinate clause was identical across conditions. Importantly, it always used the male singular pronoun (German *er*; English *he*) to refer to the subject of the main clause. This resulted in two subject gender-mismatch conditions and two subject gender-match conditions. We avoided the female singular pronoun (German *sie*; English *she*) because it is identical to the plural pronoun in German (German *sie*; English *they*), which would have resulted in resolution ambiguity. The materials further contained an interference manipulation. Although it is possible to bind the pronoun to the object in the mismatch-match condition grammatically, it cannot be resolved in this manner in the given context. Note that in all items, the subject of the main clause was the only contextually plausible referent for the pronoun.

To ensure that all children knew the names in subject position, we drew these from a ranking of children's names for 2006, the birth year of the children in our sample (Gesellschaft für Deutsche Sprache, 2006). We selected the 48 most frequent names for girls and 48 most frequent names for boys, excluding unisex names. The boys' names in the subject match conditions and the girls' names in the subject mismatch conditions were matched for length. Each group had a mean length of 5 letters ($SD = 1.7$). To ensure that all children were familiar with the direct object of the main clause, objects were chosen from the childLex corpus (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015). All objects had a normalized lemma frequency larger than 5 occurrences per million. An omnibus analysis of variance (ANOVA) showed that there was no significant difference, $F(1, 91) = 0.29$, $p = .594$, between the frequency of the male objects ($M = 43$, $SD = 54$) and that of the female objects ($M = 35$, $SD = 39$). Furthermore, the length of male objects was identical to the length of female objects, with both having a mean length of 6 letters ($SD = 2$).

Procedure

We used an EyeLink 1000 eye tracker (SR Research, Ottawa, Ontario, Canada) to record eye movements at a sampling rate of 1000 Hz. The stimulus sentences were presented on an ASUS 21-inch LCD monitor with a refresh rate of 120 Hz. Participants were seated at a monitor distance of 62 cm in a head-and-chin rest to minimize head movements. Sentences were presented using the SR Research Experiment Builder software (Version 1.10). All sentences appeared in one or two lines in Courier New font size 16. Four item lists were created to ensure that every participant read each item in only one of the four conditions. For each of the 48 items, a different sentence version was assigned to each of the four lists, and participants were assigned one of the four experimental lists in the order of their attendance.

For the children, testing took place during school hours. The paper-and-pencil tests were administered in one group session in a quiet room. The individual sessions were conducted in a separate quiet room provided by the school. Written informed consent was obtained from the children's parents, and oral consent was obtained from each child prior to testing. For the adults, testing took place in the lab facilities of the Max Planck Institute for Human Development. Adults were tested in a single session of about 60 min, with written informed consent. The study was approved by the ethics committee of the Max Planck Institute for Human Development.

A five-point calibration procedure (a moving black dot on a white backdrop) was conducted and validated until calibration accuracy reached at least 0.5° of visual angle. The calibration routine was repeated every 25 trials or after head movements. Tracking was monocular. The right eye was tracked unless tracking of the left eye considerably improved calibration. After the first calibration, the participants were presented with two practice items, each followed by a comprehension question. Participants were instructed to read the sentences silently before pushing a button on a gamepad. Upon button press, a comprehension question appeared to ensure attentive reading. The questions did not direct participants' attention to the pronoun and were designed to be answerable in all experimental conditions (e.g., "Did they run to the bank?", "Was it a tasty sandwich?"; cf. Table 1). Comprehension questions appeared randomly after 25% of trials. Along with the 48 target items, we presented 52 unrelated, structurally dissimilar fillers from a different experiment. The fillers did not contain any mismatches. After the experiment, the participants were debriefed. Those participants who did not spontaneously report detection of the mismatch during the experiment were informed about the mismatch and asked whether they had noticed it during reading. All adults had noticed the mismatch during reading, whereas roughly half of the children had noticed.

Analysis

Data were inspected and cleaned for children and adults separately using the SR Research Data Viewer software (Version 1.11.9). Fixations were cleaned automatically using the Data Viewer's implemented four-stage fixation cleaning. In a first step, short fixations with a maximum distance of 0.5° from a neighboring fixation were merged. In a second step, remaining fixations were merged with a neighboring fixation if they were shorter than 40 ms and within 1.25° distance. In a third step, all regions were checked for at least three neighboring fixations of less than 140 ms. If regions that matched this condition were found, the respective fixations were merged. In the fourth step, fixations outside a 120- to 1200-ms threshold for children, or an 80- to 1000-ms threshold for adults, were deleted from the fixation record. This step removed 7.5% of fixations of the children and 6.4% of fixations of the adults. Lastly, observations above 2.5 standard deviations from the person and item means of each dependent reading time measure were deleted for adults and children separately. For each measure, this affected roughly 2.5% of the data.

Reading time data were analyzed with linear mixed-effects models, and regression probability was analyzed with generalized linear mixed-effects models, using the lme4 package (Version 1.7) (Bates, Maechler, & Bolker, 2012) in R (R Development Core Team, 2016). Separate models were calculated for two regions: the pronoun region (always *er* and a following function word) and the sentence-final region (always a modifier and a noun). Each model included subject gender (subject gender match vs. subject gender mismatch), object gender (object gender match vs. object gender mismatch), and age (child vs. adult) as fixed effects and included participants and items as crossed random intercepts. For these regions, we calculated four reading measures: *gaze duration* (summed duration of first-pass fixations), *total reading time* (summed fixations in a region), *regression path duration* (fixations in a region, including regressions, until the region is left to the right for the first time, also called *gopast time*), and *regression probability* (percentage of saccades leaving the region to the left). All reading time measures were log-transformed to achieve a near-normal distribution. Note that the back-transformed model results are reported in milliseconds. Effect coding and Type II model comparisons were used to determine the significance of the fixed effects using the *Anova* function of the *car* package (Fox, Friendly, & Weisberg, 2013). Post hoc comparisons were estimated using cell means coding and single degree of freedom contrasts as implemented in the *glht* function from the *multcomp* package (Hothorn, Bretz, & Westfall, 2008).

Results and discussion

Global results

A 2 × 2 ANOVA with subject gender and age as crossed factors showed that comprehension accuracy was high in both adults (M = 94%, SD = 23) and children (M = 88%, SD = 33), whereas adults were more accurate on average, $F(1, 92) = 8.0$, $p < .05$. The difference in comprehension accuracy for subject gender match and subject gender mismatch was not significant, $F(1, 92) = 3.8$, $p = .054$, nor was the Subject Gender × Age interaction, $F(1, 92) = 0.7$, $p = .416$.

Analyses of eye tracking measures on the trial level showed that adults made fewer fixations (M = 19, SD = 9) than children (M = 33, SD = 19) and, therefore, had shorter mean trial reading times (M = 2941 ms, SD = 1261) than children (M = 5782 ms, SD = 2698). As a consequence, we found large main effects of age for all dependent reading time measures and so concentrate on interactions of age with subject gender and object gender, respectively. In the following paragraphs, we report results from the analyses in the two regions of interest. The observed means for the regions are reported in Table 2, and results from the linear mixed-effects models are summarized in Table 3.

Regions of interest

Pronoun region. In the pronoun region, we found an early main effect of subject gender in gaze durations for both adults and children but found no effects of object gender. Interactions with age were not significant. Pronoun regions in subject gender-mismatch sentences (M = 376 ms, SE = 15) elicited significantly longer gaze durations compared with pronoun regions in subject gender-match sentences

Table 2
Experiment 1: Observed means of eye tracking measures.

	Pronoun region				Final region			
	Subject male		Subject female		Subject male		Subject female	
Age group	Obj female	Obj male	Obj female	Obj male	Obj female	Obj male	Obj female	Obj male
				Gaze duration				
Children	612 (408)	587 (401)	647 (432)	625 (439)	918 (495)	888 (497)	879 (517)	884 (462)
Adults	290 (150)	293 (146)	309 (156)	322 (169)	427 (179)	413 (194)	432 (169)	422 (185)
				Total reading time				
Children	809 (535)	831 (570)	883 (568)	881 (553)	1100 (483)	1071 (540)	1115 (540)	1137 (527)
Adults	384 (226)	413 (233)	441 (262)	461 (274)	587 (279)	608 (271)	596 (244)	592 (254)
				Regression path duration				
Children	826 (622)	872 (685)	981 (432)	898 (672)	1268 (565)	1264 (653)	1367 (517)	1352 (610)
Adults	419 (347)	444 (364)	458 (348)	462 (359)	851 (370)	841 (353)	933 (347)	941 (400)
				Regression probability				
Children	.13 (.34)	.17 (.38)	.22 (.41)	.20 (.40)	.25 (.43)	.22 (.42)	.29 (.45)	.24 (.43)
Adults	.22 (.42)	.24 (.43)	.31 (.46)	.33 (.47)	.55 (.50)	.53 (.50)	.65 (.48)	.62 (.49)

Note. Subject male, pronoun match; Subject female, pronoun mismatch; Obj female, object female; Obj male, object male (distractor). Standard deviations are given in parentheses.

Table 3
Experiment 1: ANOVA *F* values for reading time measures and χ^2 values for regression probability.

Region	Gaze duration		Total reading time		Regression path duration		Regression probability	
	Pronoun	Final	Pronoun	Final	Pronoun	Final	Pronoun	Final
Subject	7.49**	0.30	20.26***	0.62	13.96***	15.60***	17.66***	14.99***
Object	0.23	2.55	2.91	0.01	0.00	2.27	0.64	3.09
Age	81.27***	110.27***	48.95***	74.43***	45.44***	21.02***	6.37*	33.09***
Subject × Object	0.42	3.00	0.07	0.01	2.42	0.02	0.93	0.24
Subject × Age	0.28	2.62	0.38	0.07	0.03	1.19	0.17	2.30
Object × Age	2.59	0.07	0.91	0.40	1.17	0.02	0.03	0.05
Subject × Object × Age	0.05	0.30	0.33	3.49	0.24	0.18	0.98	0.01

* $p < .05$.
** $p < .01$.
*** $p < .001$.

(M = 353 ms, SE = 15). In total reading time, there was a converging effect of subject gender in the pronoun region; mismatching pronouns took longer to read (M = 530 ms, SE = 26) than matching pronouns (M = 480 ms, SE = 24).

In regression probability, there were again main effects of subject gender but no effects of object gender and no interactions with age. In subject gender-mismatch sentences regressions were initiated in the pronoun region in 25% (SE = 2) of observations, whereas in subject gender-match sentences regressions were initiated in only 18% (SE = 2) of the observations. For regression path duration, we found a main effect of subject gender, where pronouns in subject gender-mismatch sentences elicited longer regression path durations (M = 518 ms, SE = 29) than pronouns in subject gender-match sentences (M = 467 ms, SE = 26).

In summary, the results in the pronoun region suggest that the mismatch between pronoun and subject gender disrupts bonding of pronoun to antecedent in adults and children alike. Both age groups showed longer gaze durations in the pronoun region of subject mismatch sentences than in that of subject match sentences. Furthermore, both groups showed immediate regressions out of the pronoun region when the pronoun was a mismatch to the subject antecedent. The effect of subject gender in total reading time in the pronoun region is likely a consequence of readers' regression behavior.

Final region. In the final region, there were main effects of subject gender in regressions out of the pronoun region and regression path duration, but there were no effects of object gender. Interactions with age were not significant in any of the measures. We did not find any effects of subject gender or object gender in gaze duration, suggesting no early effects in the final region. With respect to the late measures, regression path durations in the final region were significantly longer in the subject gender-mismatch sentences (M = 1066 ms, SE = 50) than in the subject gender-match sentences (M = 978 m s, SE = 46). This finding converges with the effects found in the pronoun region. The probability of regressions out of the final region was significantly higher in subject gender-mismatch sentences (M = 44%, SE = 4) than in subject gender-match sentences (M = 35%, SE = 4). We did not find any effects of subject gender or object gender in total reading times. Furthermore, there were no interactions of subject gender with object gender in any of the reported measures, nor did we find any interactions of subject gender and age. Taken together, we saw converging effects of the gender mismatch between subject and pronoun for regression probability and regression path duration, which mirrored the effects from the pronoun region. With respect to the object interference, our data do not confirm our hypotheses for the children because there were no discernible effects of object gender interference.

Experiment 2

Rationale

The results of Experiment 1 suggest that children and adults process pronouns in much the same way. This similarity was surprising because only roughly half of the children reported that they saw a pronoun mismatch in the sentences during testing when asked. We did not formally record reports of mismatch detection in Experiment 1 because the sample size would not have allowed a comparison between children who reported detecting a mismatch and those who did not. Therefore, we replicated the experiment to investigate inter-individual processing differences among the children depending on their awareness of the gender mismatch in the stimulus sentences. We used the same set of items as in Experiment 1. We were interested in two groups of children who emerged in Experiment 1: children who report mismatch detection (detectors) and those who do not report mismatch detection (non-detectors). We used reporting of the mismatch detection as an indicator of children's successful reading comprehension monitoring. We hypothesized that mismatch detection is associated with reading processing and investigated which specific reading processes are associated with successful and unsuccessful mismatch detection. We hypothesized that the detectors process the mismatching pronoun region comparably to the adults, but we expected to see differences in non-detectors. We further hypothesized that differences in the report of mismatch detection between the detectors and non-detectors are associated with individual differences in the component skills of reading. Therefore, we compared differences in reading comprehension, reading fluency, and efficiency of auditory sentence comprehension between the two groups.

Method

Participants

The sample in Experiment 2 was a subset of the Berlin Developmental Eye Tracking Study (DevTrack), for which a total of 92 children from two primary schools in Berlin were tested. Of these, 75 successfully completed the experimental sessions for the current study. Of the participating children, 2 were early immigrants who arrived in Germany before 5 years of age. All children received their reading instruction in German only. Of the participating children, 41 were girls. The study was conducted in the winter term of the children's fourth school year, when they were 9 years old (*M* = 11 9 months, *SD* = 6). All children reported normal or corrected-to-normal vision.

Materials

The sentence materials for the eye tracking session were identical to those in Experiment 1 (see Table 1). We again assessed reading comprehension with the ELFE 1–6 (Lenhard & Schneider, 2006) and assessed reading fluency with a subtest of the SLRT-II (Moll & Landerl, 2010). Furthermore, we tested children's auditory syntactic integration skill using the computerized ProDi-L (Richter, Isberner, Naumann, & Kutzner, 2012).

Reading comprehension. The German reading comprehension test ELFE 1–6 contains three subtests that target word comprehension, sentence comprehension, and text comprehension. In the word comprehension subtest, children receive a list of pictures and need to mark the word for the depicted item from a list of five given words. In the sentence comprehension subtest, children are asked to insert the appropriate word in a sentence context from a choice of five words. The ELFE 1–6 text comprehension subtest comprises a list of short texts, each of which has a set of questions tapping various levels of comprehension.

Reading fluency. The SLRT-II reading fluency subtest contains a list of words that need to be read out loud. Every child is given 1 min to read as many words as possible until the test terminates. The fluency score is calculated by the number of words read, corrected for misses and omissions.

Syntactic integration. An auditory version of the syntactic integration subtest of the German ProDi-L was used to assess the efficiency of syntactic integration. Children listened to a list of 40 sentences via headphones and were asked to press a green button when the sentence was correct and a red button when the sentence was incorrect. Half of the sentences contained morpho-syntactic errors such as a wrong word order and faulty case marking. Children's response accuracy was analyzed along with reaction time, the latter as a marker for the efficiency of morpho-syntactic processing.

Procedure

The participating children completed the ELFE 1–6 reading comprehension test as part of a group session in their classroom, and they completed the SLRT-II and ProDi-L computerized tests in separate individual sessions. The eye tracking setup was the same as in Experiment 1 except that for technical reasons the presentation software was exchanged for the University of Massachusetts' EyeTrack (Version 7.10) (Stracuzzi & Kinsey, 2006). The sentences appeared in one or two lines in the middle of the screen in a monospaced font (Courier New size 14). Upon arrival, the children were assigned to one of the four stimulus lists by order of appearance and were asked for oral consent prior to testing. The procedure was the same as in Experiment 1.

In Experiment 2, we also assessed mismatch detection during testing by asking children a series of questions after the first block (approximately 20 trials). If a child had not reported the mismatch spontaneously by the first pause, the experimenter would first ask whether the child had noticed something weird in the sentences. If the response was negative, the experimenter continued to ask, "You know, sometimes one does not understand a word during reading or a word seems wrong in the sentence. Did that happen in what you have just read?" If the response remained negative, the experimenter would just make an encouraging remark ("You're doing a really good job. Ready to move on?") and continue to the next block. If a child reported the mismatch spontaneously or after inquiry, the experimenter would say, "You spotted a mistake, good job! That may happen again, but you can just continue reading quietly." The children were not prompted again during the experiment, but if they reported the gender mismatch anywhere during the remainder of the session, this was rated as positive report of mismatch detection.

Analysis

Data were cleaned as follows. First, fixations of less than 80 ms were combined with a neighboring fixation if it was within one character. Short fixations of 40 ms or less were deleted if they occurred within three characters of a neighboring fixation. Second, only fixations within a threshold of 120–1200 ms were kept for analysis. This cleaning procedure removed less than 2% of fixations of each measure. Before models were calculated, observations above 2.5 standard deviations from the person or item mean of each dependent measure were removed. This removed less than 3% of the data in each measure.

The cleaned data were transformed and analyzed using the same methods and dependent variables as in Experiment 1 except that in Experiment 2 the models included incongruence detection (detector vs. non-detector) as a fixed effect. We further used Welch's two-sample t test to compare reading comprehension, reading fluency, and auditory sentence comprehension of the detector and non-detector groups.

Results and discussion

In Experiment 2, we replicated our main findings for the children from Experiment 1. The observed means are summarized in Table 4, and the effects from the linear mixed-effects models are summarized in Table 5.

Our inquiry during testing resulted in 43 children (57%) who reported mismatch detection and 32 children who never reported mismatch detection. Children who did not report the mismatch, however, may still have been aware of it on some level without explicitly verbalizing it. For simplicity, we refer to the children who reported mismatch detection as detectors and the children who did not report mismatch detection as non-detectors, but it is important to keep in mind that not reporting a mismatch does not necessarily presuppose absence of awareness.

Table 4
Experiment 2: Observed means of eye tracking measures.

Group	Pronoun region				Final region			
	Subject male		Subject female		Subject male		Subject female	
	Obj female	Obj male	Obj female	Obj male	Obj female	Obj male	Obj female	Obj male
	Gaze duration							
Non-detectors	438 (241)	407 (233)	474 (287)	487 (312)	724 (385)	754 (401)	699 (395)	720 (399)
Detectors	381 (219)	391 (213)	431 (268)	421 (251)	736 (410)	731 (407)	651 (379)	677 (399)
	Total reading time							
Non-detectors	603 (358)	561 (358)	672 (422)	667 (418)	976 (501)	961 (482)	967 (491)	1017 (515)
Detectors	631 (420)	632 (369)	790 (477)	784 (499)	1061 (534)	1051 (505)	1031 (510)	1049 (534)
	Regression path duration							
Non-detectors	538 (413)	526 (401)	601 (462)	582 (444)	1204 (718)	1201 (710)	1288 (738)	1330 (763)
Detectors	488 (404)	493 (415)	582 (494)	554 (442)	1501 (760)	1549 (770)	1646 (812)	1616 (796)
	Regression probability							
Non-detectors	.13 (.34)	.15 (.36)	.17 (.38)	.15 (.38)	.37 (.48)	.37 (.48)	.42 (.49)	.45 (.50)
Detectors	.23 (.42)	.21 (.41)	.30 (.46)	.34 (.47)	.65 (.48)	.65 (.48)	.73 (.45)	.73 (.44)

Note. Subject male, pronoun match; Subject female, pronoun mismatch; Obj female, object female; Obj male, object male (distractor). Standard deviations are given in parentheses.

Table 5
Experiment 2: ANOVA *F* values for reading time measures and χ^2 values for regression probability.

Region	Gaze duration		Total reading time		Regression path duration		Regression probability	
	Pronoun	Final	Pronoun	Final	Pronoun	Final	Pronoun	Final
Subject	29.25***	19.40***	103.71***	0.04	32.57***	22.93***	23.93***	27.08***
Object	0.94	1.57	1.25	0.30	1.19	0.22	0.10	0.34
Detection	1.54	0.94	2.83	1.01	0.79	16.93***	15.71***	31.64***
Subject × Object	0.00	0.73	0.07	2.73	0.41	0.00	0.58	0.03
Subject × Detection	1.13	1.58	6.84**	1.81	0.79	0.26	5.00*	0.64
Object × Detection	1.32	0.04	1.34	0.10	0.41	0.09	0.02	0.12
Subject × Object × Detection	3.39	0.00	1.66	0.59	0.41	1.18	3.99*	0.07

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Global results

Mean accuracy in the comprehension questions was high (M = 93%, SD = 26). An ANOVA over response accuracy with subject gender and detection as crossed factors showed that accuracy was not affected by subject gender match, $F(1, 146) < 1$, $p = .588$. There was a main effect of detection, such that accuracy was slightly higher in the detectors (M = 94%, SD = 10) than in the non-detectors (M = 90%, SD = 13), $F(1, 146) = 4.23$, $p < .05$. There was no interaction of subject gender and detection, $F(1, 146) < 1$, $p = .674$.

The results of the analyses of individual differences for the two detector groups are summarized in Table 6.

Detectors and non-detectors did not differ in reading comprehension skill, $t(62) = -1.8$, $p = .078$. Our analyses of accuracy in the sentence processing task showed that both groups correctly rejected

Table 6
Analyses of group differences.

Measure	Detectors		Non-detectors	
	M	*SD*	*M*	*SD*
Reading comprehension	90.3	12.8	84.5	15.0
Reading fluency	83.8	17.6	73.2	14.4
Sentence comprehension accuracy	.88	.07	.88	.07
Sentence comprehension reaction time	3011	323	2896	232

Note. Raw scores for the reading comprehension test, raw scores for the reading fluency test, and accuracy and reaction time in the sentence comprehension test are shown.

sentences with morpho-syntactic errors, such that there was no difference between the groups for accuracy, $t < 1$, $p = .968$. Furthermore, we did not find a significant difference for the efficiency of auditory processing between the two groups, $t = 1.8$, $p = .075$. However, the two groups differed in reading fluency, with the detectors being the more fluent readers, $t(74) = -2.9$, $p = .005$. In the reading fluency task, the detectors read significantly faster than the non-detectors (about 10 words more per minute). The results of the reading fluency test concur with our observations in the eye tracking measures. Analyses of eye tracking measures on the trial level showed that children who reported detection of the mismatch spent less time on first-pass reading than children who did not report the mismatch. The observed mean gaze duration per region for the detectors ($M = 480$ ms, $SD = 121$) was significantly shorter than that for the non-detectors ($M = 552$ ms, $SD = 129$), $t(65) = 2.5$, $p = .016$. However, total reading times did not differ between the detectors and non-detectors. The detectors spent as much time reading each region ($M = 786$ ms, $SD = 224$) as the non-detectors ($M = 766$ ms, $SD = 216$), $t(69) < 1$, $p = .708$.

In summary, eye movement measures on the trial level revealed that the detectors read faster during the first pass than the non-detectors. However, there was no difference between the groups in total reading time, from which we may conclude that the detectors spent more time rereading the sentences. The non-detectors made regressions less frequently, and their gaze durations were correlated more closely with their total reading times.

Regions of interest

Pronoun region. In the pronoun region, there was a main effect of subject gender in gaze durations, such that subject gender-mismatch sentences elicited significantly longer gaze durations ($M = 379$ ms, $SE = 14$) than subject gender-match sentences ($M = 347$ ms, $SE = 14$). For total reading time, we also found a significant effect of subject gender in the pronoun region. The region was read longer in subject gender-mismatch sentences ($M = 594$ ms, $SE = 25$) than in subject gender-match sentences ($M = 497$ ms, $SE = 21$). Detectors showed a significant 131-ms effect of subject gender, which was significantly larger than the 67-ms effect of subject gender in non-detectors, $t = 2.61$, $p = .009$. A converging effect emerged in regression path duration. In the subject gender-mismatch sentences, children took significantly longer to pass the pronoun region ($M = 452$ ms, $SE = 19$) than to pass the same region in the subject gender-match sentences ($M = 406$ ms, $SE = 17$). For regression probability, we again found a main effect of subject gender, such that children were more likely to initiate regressions from the pronoun region in the subject gender-mismatch sentences ($M = 21\%$, $SE = 2$) than from the pronoun region in the subject gender-match sentences ($M = 15\%$, $SE = 2$). Furthermore, there was a main effect of detection, such that, in general, the detectors were more likely to make regressions out of the pronoun region ($M = 24\%$, $SE = 2$) than the non-detectors ($M = 13\%$, $SE = 2$). Post hoc contrasts revealed that in fact, the subject gender effect in the pronoun region was entirely driven by the detectors. There was a significant simple main effect of subject gender in the detector group, $t = 5.4$, $p < .001$, but not in the non-detector group, $t = 1.1$, $p = .268$. For the detectors, the likelihood to make a regression out of the pronoun region in subject gender-mismatch sentences was higher ($M = 30\%$, $SE = 3$) compared with subject gender-match sentences ($M = 19\%$, $SE = 2$), whereas the non-detectors made equally few regressions out of the pronoun region in subject gender-match sentences ($M = 14\%$, $SE = 2$) and subject gender-match sentences ($M = 12\%$, $SE = 2$).

Lastly, we found a three-way interaction of incongruence detection, subject gender, and object gender in the pronoun region. The detectors made numerically more regressions out of the pronoun region of subject gender-mismatch sentences in the object gender-match condition than in the object gender-mismatch condition. The difference, however, was nonsignificant, $t = -1.83$, $p = .067$. The non-detectors, in contrast, made numerically more regressions in the subject gender mismatch sentences when the object was a mismatch as well, although this difference was also nonsignificant, $t = 1.18$. $p = .237$.

In summary, the results in the pronoun region suggest that children use pronoun gender for rapid pronoun–antecedent bonding during online reading given that the reading processing of both groups of children was disrupted immediately in the subject gender-mismatch conditions. Our analyses of the eye movement behavior of detectors and non-detectors revealed important differences in late measures, such that the detectors make more regressions. These resulted in longer total reading times in the pronoun region for the detectors compared with the non-detectors.

Final region. In the final region, there were main effects of subject gender and main effects of detection but no interactions. There were no effects of object gender. Gaze durations in the final region showed an early effect of subject gender in the opposite direction from the pronoun region. Final regions of sentences with a *matching* subject elicited significantly longer gaze durations ($M = 628$ ms, $SE = 24$) than final regions of sentences with a *mismatching* subject ($M = 578$ ms, $SE = 22$). Note that in Experiment 1 we did not find any effect of subject for gaze duration in the final region. There were no effects for total reading time in the final region, consistent with our findings in Experiment 1.

The delayed effects of subject gender went in the same direction as in the pronoun region. There were prolonged regression path durations for the final region of subject mismatch sentences ($M = 1\,286$ ms, $SE = 54$) compared with subject match sentences ($M = 1183$ ms, $SE = 49$ ms). Moreover, the detector group showed significantly longer regression path durations in the final region ($M = 1428$ ms, $SE = 72$ ms) compared with the non-detector group ($M = 1066$ ms, $SE = 60$). This result converges with the longer trial reading times of the detectors and their regression behavior; for regression probability, we found that subject mismatch sentences elicited more regressions in general ($M = 61\%$, $SE = 3$) than subject match sentences ($M = 51\%$, $SE = 3$). We also found a large main effect of incongruence detection in the final region, such that the detectors made more regressions out of this region ($M = 73\%$, $SE = 4$) than the non-detectors ($M = 38\%$, $SE = 4$). Notably, there was no interaction of subject and incongruence detection in the final region, suggesting that both groups show sentence-final mismatch effects.

In summary, Experiment 2 revealed significant differences in the processing of children who report mismatch detection and children who do not report it. The fact that the detectors made more regressions overall explains the disparity of gaze duration and total reading time between the two groups in the regions of interest. With respect to the object gender manipulation, we replicated our null effect from Experiment 1. Lastly, we saw a novel effect in the final region that we did not obtain in Experiment 1. In the final region, there were prolonged gaze durations for subject gender-match sentences compared with subject gender-mismatch sentences. We discuss this finding in more detail in the General Discussion.

General discussion

The main focus of the current study was on children's sensitivity to gender feature mismatches during the reading of pronouns and their susceptibility to interference of feature-matching entities during reading. Our materials consisted of two-clause sentences with a male pronoun in the second clause such as "Leon/Lisa shooed away the sparrow/the seagull and then he ate the tasty sandwich." The pronoun always referred to the subject of the main clause, which was either a gender-matching or gender-mismatching name (Leon{m}/Lisa{f}). Furthermore, the sentences contained an interfering direct object that was either a gender match or mismatch to the pronoun (the sparrow{m}/the seagull{f}). We recorded children's and adults' eye movements while they read the sentences, focusing our analyses on the pronoun and sentence-final regions, the latter of which directly followed the pronoun region. In Experiment 1, results suggested no qualitative differences between children's and adults' processing of the pronouns. Both age groups showed immediate sensitivity to the subject gender mismatch and no effects of interference from the object. However, in contrast to the adults, not all children seemed to detect the subject gender mismatch. In Experiment 2, we replicated our finding from Experiment 1 with a larger sample of children and found that 43% of the children reported that they were unable to detect the mismatch during the experiment. Although those children who detected the gender mismatch processed the mismatching pronoun comparably to the adults, there were important differences in pronoun processing between the "detectors" (i.e., children who did report the mismatch when prompted) and the "non-detectors" (i.e., children who did not report the mismatch).

Processing of the pronoun region in children and adults

In our experiments, we aimed to tap early and late reading processes using different eye tracking measures: gaze duration for early stages of processing and total reading time, regression behavior, and regression path duration for later stages of processing. We hypothesized that children's processing efficiency may differ from adults' and that the eye movement measures should be affected by these processing differences. In summary, our results suggest that pronoun–antecedent association based on gender cue is indeed as automatic in children as in adults (Rigalleau et al., 2004). This was reflected in the longer gaze durations on the pronoun for mismatching subjects than for matching subjects. In Experiment 1, we did not find any differences between children and adults with respect to gaze durations. Likewise, in Experiment 2, both children who reported mismatch detection and children who did not report it showed longer gaze durations for gender-mismatching pronouns than for gender-matching pronouns. Whether children resolve a pronoun successfully, however, seems to be subject to individual processing differences.

Both groups of children slowed down during the first pass of the region when a mismatching pronoun was presented. This suggests that gender information was automatically registered as incongruent even by those children who were unable to report a mismatch after reading. We assume that children who reported the mismatch also understood the correct resolution of the pronoun. It is important to note that, in contrast, absence of reporting does not necessarily imply absence of comprehension by the non-detectors. Therefore, the group of non-detectors is defined less clearly. Our results from gaze durations indeed show that they too are sensitive to the gender mismatch. We did not ask comprehension questions tapping pronoun resolution in this study because such questions would have prompted the participants to explicitly pay attention to pronoun inconsistencies. This would likely have interfered with natural reading. Successful monitoring and reading comprehension, however, have been shown to be closely related in children (van der Schoot, Reijntjes, & van Lieshout, 2012).

Although both children who reported the mismatch and children who did not report the mismatch slowed down during gaze duration at a mismatching pronoun compared with a matching pronoun, the detectors were more likely to reread the mismatching area. Successful integration, as evidenced by the report of mismatch detection, was associated with higher regression probability for mismatching pronouns than for matching pronouns in the pronoun region. This is what led to longer total reading times of the mismatching pronoun. We interpret this finding in terms of processing depth of the pronoun. As originally observed by Rayner (1998), readers may make very short regressions (up to one word to the left) when the currently fixated word disrupts fluent text processing. It has been suggested that in this way readers may delay new input from upcoming words in order to allow for more processing time in a conflictive sentence region. The eye movement pattern we found in the adults and detectors may constitute a "coping mechanism" in the face of local processing difficulty, leading to an increase in processing depth. Thus, more processing time in the critical region—that is, directly at the pronoun—is associated with mismatch detection by way of deeper processing of the pronoun.

The subject gender mismatch was also evident in regression probability from the sentence-final region, where we saw a main effect of subject gender and of mismatch detection. The detectors made more regressions overall at the end of a sentence. Note that this matches the adult reading behavior. The result is also in line with earlier findings connecting rereading probability and pronoun comprehension (Ehrlich et al., 1999). Our results show that whereas immediate regressions are associated with the reporting of mismatch detection in children, the detectors also make more regressions overall. Arguably, monitoring of comprehension is closely connected to rereading behavior in children.

Individual differences in children's reading processing

The eye movement behavior associated with successful comprehension monitoring in the detectors can be described as faster first-pass reading, evident as shorter gaze durations, combined with more regressive saccades in the mismatching pronoun region compared with the non-detectors. In addition to the processing data, we tested children's reading comprehension, reading fluency, and syntactic integration skill. Contrary to our expectations, we did not find any differences in the non-detectors' and detectors' reading comprehension skill or in their accuracy or efficiency of syntactic integration. Although the striking similarities between the detectors' and adults' reading processing may suggest that the non-detectors are less developed readers, we would expect significant differences in the reading comprehension test if this were the case. Although reading comprehension and syntactic integration were comparable between the groups, the non-detectors were significantly slower in the reading fluency test. This result concurs with our processing data, which showed that the detectors had shorter gaze durations than the non-detectors at the trial level. We may conclude that processing speed is a main determinant of reading comprehension and monitoring in children. The association of slow decoding speed and comprehension monitoring failure in our study is in line with the lexical quality hypothesis (Perfetti, 1994; Perfetti & Hart, 2001). The lexical quality hypothesis assumes that slow word reading is a sign of inefficient and effortful extraction of word information such as orthographic representation and semantic meaning. If this process is effortful, children may lack the cognitive resources to engage in monitoring of comprehension. Therefore, we conclude that children who read more fluently can make the necessary resources available for comprehension monitoring and, as a result, can report the mismatch when prompted.

The pattern of our findings is in line with a prior study on children's detection of semantic anomalies. Connor et al. (2014) studied fifth graders' eye movements in two-sentence stories that contained local semantic anomalies, for example, "truck" when "plane" would be appropriate: "Last week Kyle *flew* to visit his family in another city. The large *plane*/*truck* was spacious and quickly transported him." The authors reported that children with stronger literacy skill read more fluently and made more regressions at the semantic anomaly than children with weaker literacy skill, although both groups' first-pass reading was slowed down by the anomaly. The authors concluded that slowing down during first-pass reading at a semantic incongruence in the text is largely automatic and that successful comprehension is determined by the extent of a child's engagement with the text after a semantic incongruence was detected. Our results show that this pattern transfers to pronoun processing. The mismatching pronoun induced a local processing disruption given that children who reported the mismatch made regressions immediately. In addition, there was a reverse effect of subject gender mismatch for gaze durations in the final region. Contrary to our hypotheses, we found prolonged gaze durations for matching pronouns rather than mismatching pronouns. We interpret this finding in terms of a wrap-up for congruent sentences but not for incongruent sentences. It is conceivable that children do not attempt meaning integration at the sentence level for incongruent sentences at all. This interpretation of our results suggests that even surface-level incongruences can lead to unresolved sentence processing in children. Because we did not see this reverse effect of subject gender mismatch in the adults, we would hypothesize that the processing disruption induced by a mismatching pronoun has a more lasting effect on less proficient readers than on proficient readers. It has previously been shown that incongruent pronouns affect reaction time to comprehension questions, but not comprehension accuracy, in adults (Rigalleau et al., 2004). This effect may be individually different in developing readers, who might not recover from conflicting surface information in text as quickly as proficient readers.

Null effect of object gender

Lastly, we turn to the null effect of the object gender manipulation. We did not obtain any main effects of object gender match for pronoun processing in our experiment. In contrast to our hypotheses, the intervening object did not influence processing of the matching pronoun regions or the mismatching pronoun regions. Findings for the online reading pattern associated with feature-based interference are mixed in the literature (Jäger et al., 2017), and our null results do not exclude the existence of cue-based interference effects during pronoun resolution in children. We suggest that future research pursuing such interference effects use paradigms more closely related to those used with adults (e.g., Badecker & Straub, 2002; Cunnings & Felser, 2013; Patil, Vasishth, & Lewis, 2016).

Conclusion

Our experiments show that examining local processing strategies at key points in a sentence may inform our understanding of children's reading comprehension and potential sources of difficulties. The results from Experiment 1 suggest no developmental differences between children's and adults' pronoun processing. This indicates that there are no overall qualitative differences in pronoun processing between beginning and proficient readers. Experiment 2, however, showed that there are inter-individual differences in children's processing of mismatching pronouns. These processing differences were associated with children's reporting of the mismatch when prompted. Importantly, total reading time was not itself related to mismatch detection; rather, those children who reported detection of the mismatch allocated additional time to rereading only. Therefore, we may also conclude that total time on task during reading is not a sufficient indicator of reading performance. This is important because educators tend to assess children's reading performance in classroom settings based on how much time the children need to finish reading a passage. To further understand how pronoun resolution influences children's text comprehension at large, we suggest that whole text passages be taken into account. Because pronouns serve as an anchor for textual coherence, their resolution is important for text comprehension. Our findings using tightly experimentally controlled sentences, therefore, are likely to transfer to less constrained natural text reading.

Acknowledgment

We thank Sophia Tischer for assistance in data collection and preprocessing.

References

Ariel, M. (2004). Accessibility marking: Discourse functions, discourse profiles, and processing cues. *Discourse Processes, 37*, 91–116.

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

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*, 748–769.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package Version 0.999375-42. <http://cran.R-project.org/package=lme4>.

Blythe, H. I., & Joseph, H. (2011). Children's eye movements during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *Oxford handbook on eye movements* (pp. 643–662). Oxford, UK: Oxford University Press.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*, 31–42.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology, 49*, 639–663.

Connor, C. M. D., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2014). Individual differences in fifth graders' reading and language predict their comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading, 19*, 114–134.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes, 28*, 188–219.

Ehrlich, M. F., Rémond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing, 11*, 29–63.

Eilers, S., Tiffin-Richards, S., & Schroeder, S. (2018). The repeated name penalty effect in children's natural reading. Evidence from eye tracking. *Quarterly Journal of Experimental Psychology*. doi: 0.1177/1747021818757712.

Elbro, C., Oakhill, J., Megherbi, H., & Seigneuric, A. (2017). Aspects of pronominal resolution as markers of reading comprehension: The role of antecedent variability. *Reading and Writing, 30*, 813–827.

Fagan, S. M. (2009). *German: A linguistic introduction*. Cambridge, UK: Cambridge University Press.

Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the *car* package. *The R Journal, 5* (1), 39–52.

Gagl, B., Hawelka, S., & Wimmer, H. (2015). On sources of the word length effect in young readers. *Scientific Studies of Reading, 19*, 289–306.
Garnham, A., & Oakhill, J. (1985). On-line resolution of anaphoric pronouns: Effects of inference making and verb semantics. *British Journal of Psychology, 76*, 385–393.
Garnham, A., Oakhill, J., & Johnson-Laird, P. N. (1982). Referential continuity and the coherence of discourse. *Cognition, 11*, 29–46.
Gesellschaft für Deutsche Sprache (Ed.). (2006). *Liste der beliebtesten Vornamen*. Wiesbaden, Germany: Author.
Hickmann, M., Schimke, S., & Colonna, S. (2015). From early to late mastery of reference. In L. Serratrice & S. E. M. Allen (Eds.). *The acquisition of reference* (pp. 181–211). Amsterdam: John Benjamins.
Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346–363.
Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language, 94*, 316–339.
Joseph, H. S., Bremner, G., Liversedge, S. P., & Nation, K. (2015). Working memory, reading ability, and the effects of distance and typicality on anaphor resolution in children. *Journal of Cognitive Psychology, 27*, 622–639.
Joseph, H. S., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 61*, 708–723.
Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: Center for the Study of Language and Information.
Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*, 1–44.
Lenhard, W., & Schneider, W. (2006). *ELFE 1–6: Ein Leseverständnistest für Erst-bis Sechstklässler*. Göttingen, Germany: Hogrefe.
Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest: Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)*. Göttingen, Germany: Hogrefe.
Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *Quarterly Journal of Experimental Psychology, 40*, 693–718.
Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39*, 85–101.
Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, UK: Blackwell.
Oakhill, J. V. (1982). Constructive processes in skilled and less-skilled comprehenders. *British Journal of Psychology, 73*, 13–20.
Oakhill, J. V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology, 54*, 31–39.
Oakhill, J., & Yuill, N. (1986). Pronoun resolution in skilled and less-skilled comprehenders: Effects of memory load and inferential complexity. *Language and Speech, 29*, 25–37.
Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00329.
Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed). *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86). Washington, DC: American Psychological Association.
Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). Amsterdam: Elsevier.
R Development Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.
Richter, T., Isberner, M.-B., Naumann, J., & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern [Process-based measurement of reading skills in primary school children]. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology, 26*, 313–331.
Rigalleau, F., Caplan, D., & Baudiffier, V. (2004). New arguments in favour of an automatic gender pronominal process. *Quarterly Journal of Experimental Psychology A, 57*, 893–933.
Schroeder, S., Hyönä, J., & Liversedge, S. P. (2015). Developmental eye-tracking research in reading: Introduction to the special issue. *Journal of Cognitive Psychology, 27*, 500–510.
Schroeder, S., Würzner, K. M., Heister, J., Geyken, A., & Kliegl, R. (2015). ChildLex: A lexical database of German read by children. *Behavior Research Methods, 47*, 1085–1094.
Sekerina, I. A. (2015). Online evidence for children's interpretation of personal pronouns. In L. Serratrice & S. E. M. Allen (Eds.), *The acquisition of reference* (pp. 213–240). Amsterdam: John Benjamins.
Stracuzzi, D., & Kinsey, J. (2006). EyeTrack (Version 0.7.10). Amherst: University of Massachusetts.
van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor comprehenders. *Reading and Writing, 25*, 1665–1690.
Yuill, N., & Oakhill, J. (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology, 79*, 173–186.
Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.