

ARTICLE

Print exposure predicts pronoun comprehension strategies in children

Jennifer E. ARNOLD^{1*}, Laura CASTRO-SCHILO², Sandra ZERKLE¹, and Leela RAO¹

¹Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill and

²SAS Institute Inc.

*Corresponding author: UNC Chapel Hill, Dept. of Psychology and Neuroscience, Davie Hall #337B, CB #3270, Chapel Hill, NC 27599-3270. E-mail: jarnold@email.unc.edu

(Received 8 July 2018; revised 27 November 2018; accepted 18 February 2019;
first published online 24 May 2019)

Abstract

Language development requires children to learn how to understand ambiguous pronouns, as in *Panda Bear is having lunch with Puppy. He wants a pepperoni slice*. Adults tend to link *he* with Panda Bear, the prior grammatical subject, but young children either fail to exhibit this bias (Arnold, Brown-Schmidt & Trueswell, 2007) or do so more slowly than adults (Hartshorne *et al.*, 2015a; Song & Fisher, 2005). In the current study, we test whether language exposure affects this bias in elementary-school-age children. Children listened to stories like the one above, and answered questions like “Who wants a pepperoni slice?” which reveal their pronoun interpretation. Individual variation in the rate of selecting the subject character correlated with measures of print exposure, such that children who read more are more likely to follow the subject bias. This is the first study to establish that print exposure affects spoken pronoun comprehension in children.

Keywords: reading print exposure; pronoun comprehension

When children learn a language, they have to learn more than just words and grammar. Children also must learn how to interpret language appropriately in context, for example drawing the correct inferences to interpret ambiguous personal pronouns like *she* and *he*. Pronouns are highly ambiguous, and yet they are one of the most frequent words in the language, and somehow people understand each other. Adults rely on syntactic structure, semantics, and social cues, associating pronouns with the most cognitively salient referent in the context (Arnold, 1998; Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000; Chafe, 1976; Gernsbacher & Hargreaves, 1988; Gundel, Hedberg, & Zacharski, 1993). When more than one referent has been recently mentioned, adults are biased toward the grammatical subject, which is usually the first-mentioned entity in English. For example, in (1) the pronoun *she* refers to Mother Bat, and not Stellaluna.

- (1) Each night, Mother Bat would carry Stellaluna clutched to her breast as she flew out to search for food. (from *Stellaluna*, by Janell Cannon, 1993)

The subject bias is not categorical, and indeed pronouns sometimes refer to non-subjects, as in example (2).

- (2) One day George saw a man. He had on a large yellow hat. (from *Curious George* by H. A. Rey, 1969).

Nevertheless, this bias is robust for adults, and has emerged in studies using multiple methodologies (e.g., Arnold *et al.*, 2000; Crawley & Stevenson, 1990; Gordon, Grosz, & Gilliom, 1993; Järvikivi, van Gompel, Hyönä, & Bertram, 2005; Kaiser, 2011; McDonald & MacWhinney, 1995; van Rij, van Rijn, & Hendriks, 2011, 2013).

Where does this linguistic bias come from? Our study tests the role of language exposure, namely the idea that children develop the ability to follow the subject bias through exposure to language. We hypothesize that language exposure is important because it illustrates the types of discourse patterns that are most frequent. These patterns might have to do with how people refer overall (who is going to be mentioned next?) or they might have to do with the behavior of pronouns (what kinds of things do pronouns usually refer to?).

Our specific goal is to demonstrate that individual differences in language exposure predict performance on a spoken pronoun comprehension task. While we assume that all types of exposure matter, here we test individual differences in one type of exposure, namely exposure to books. Print exposure is a useful metric because we can measure it more easily than variation in spoken language input. In addition, we hypothesize that literate language may provide an especially helpful domain for learning discourse patterns.

We examine the effect of exposure on performance in a spoken language comprehension task, where children could potentially interpret pronouns either as a function of the linguistic subject bias, or as a function of the speaker's gaze. We know that adults tend to follow social cues like pointing and gazing (Goodrich Smith & Hudson Kam, 2015; Nappa & Arnold, 2014). However, relatively less is known about how social cues guide pronoun comprehension for children. Therefore our secondary research question is whether children also use gaze to guide pronoun interpretation.

Before describing our study, we discuss why the question of exposure is theoretically relevant, how it relates to previous findings with adults and children, and why it is worth examining the role of gaze.

Does language exposure guide the development of pronoun comprehension strategies?

Our study tests what we call the 'exposure hypothesis', namely the idea that people learn discourse conventions through exposure to language. We focus specifically on the subject bias, which illustrates the tendency for adults to link pronouns with prominent referents (Almor, 1999; Ariel, 1990; Arnold, 1998, 2010; Arnold *et al.*, 2000; Chafe, 1976; Foraker & McElree, 2007), specifically the grammatical subject of the preceding clause. On the exposure hypothesis (Arnold *et al.*, 2007), this bias is learned through exposure to discourses, which illustrate the most common patterns of reference. This view contrasts with the 'natural prominence' position (Song & Fisher, 2005), which suggests that the prominence of subjects is learned early because subjects encode prominent semantic roles (like agent), and things that occur early in the sentence tend to have a memory advantage.

We review four reasons to expect that language exposure might influence the development of pronoun comprehension. First, current theories of reference comprehension suggest that listeners calculate referential probabilities, which may be based on experience with referential frequencies. Second, experiments with young children show that they do not follow the subject bias as consistently or as quickly as adults, raising questions about how this skill develops. Third, experiments with adults demonstrate that individual differences in print exposure account for variability in pronoun comprehension. Fourth, evidence suggests that language exposure affects other types of language development.

Probability-driven theories of reference comprehension

Current models of reference comprehension suggest that listeners interpret reference by calculating referential probabilities (e.g., Arnold, 1998; Frank & Goodman, 2012; Hartshorne, O'Donnell, & Tenenbaum, 2015a; Kehler, Kertz, Rohde, & Elman, 2008; Kehler & Rohde, 2013; Rohde & Kehler, 2014). A central component to all these models is the representation of $p(r)$, namely the probability that a referent will be mentioned. Note that this probability is independent of form. That is, this is not a question of what pronouns tend to refer to, but rather who a speaker is likely to refer to at all. If listeners track referential probability, it would allow them to anticipate upcoming references before they even encounter the referential expression (e.g., a pronoun, name, or description). In the case of ambiguous forms, this anticipation helps listeners identify the referent.

In addition, some models suggest that listeners also keep track of the pronoun-specific probabilities. For example, Kehler *et al.* (2008; see also Kehler & Rohde, 2013; Rohde & Kehler, 2014) propose a model that contains a representation of the likelihood that a speaker would use a pronoun for a particular referent, which would require tracking the frequency of pronouns referring to particular types of referents, e.g., subjects. While their computational model is not a model of development, it might suggest that children track the types of things that pronouns refer to specifically, independent of other forms of reference.

Under this theoretical framework, a central question is how children learn to calculate referential probabilities. One hypothesis is that they learn these probabilities by observing the most frequent discourse patterns (Arnold, 1998; 2010; Arnold *et al.*, 2007; Arnold, Strangmann, Hwang, Zerkle, & Nappa, 2018a; Arnold, Strangmann, Hwang, & Zerkle, 2018b). For example, if speakers and writers tend to refer to subjects frequently, children might learn that, for a given referent that occurs in subject position, there is a high probability of future reference to this entity.

Indeed, evidence shows that reference to subjects is quite frequent. Arnold (1998) analyzed texts from children's books. When writers re-mentioned an entity from the previous clause ($n = 240$) they were far more likely to refer to the subject (74%) than an object or oblique (see also Arnold *et al.*, 2018b, for a more stringent re-analysis of these data). Arnold (2001) analyzed a sample of utterances with transfer verbs (e.g., *give*, *take*) from the Aligned-Hansard corpus (which contains transcripts from Canadian parliamentary proceedings), and asked whether the first referent in the next utterance referred to the subject, prepositional object, or something else. She found that about 40% of first references referred to the subject, but only about 8% referred to the prepositional object.

These corpus analyses suggest that subject reference is highly frequent for both spoken and written language. One possibility is that people learn to associate

grammatical subject position with a high probability of re-mention (Arnold, 1998; Arnold *et al.*, 2007). If so, learning this pattern may be facilitated by more frequent exposure to linguistic discourses, or perhaps exposure to more complex discourses. For example, to learn the subject bias, people may need to be exposed to utterances with more than one referent in order to notice that subjects and objects/obliques are different in how often they are mentioned. While children may hear multiple-referent utterances in spoken language, it seems plausible that this sort of context may be more common in written language. If so, exposure to print materials may be an especially robust domain for learning the subject bias.

Alternatively, language exposure may be important for learning the specific behavior of pronouns (cf. Kehler *et al.*, 2008). If so, exposure may be important for providing experience with how pronouns behave. This experience may be especially relevant for third person pronouns (*he, she, it, they*), which depend more on the linguistic context than first and second person pronouns (*I, you, we*), which depend on knowing who the speaker and addressee are. Written language is more socially decontextualized than spoken language, which may make it a good domain for learning about third person pronouns.

An additional possibility is that language exposure, especially exposure to genres like written language, is important for learning very general strategies like ‘pay attention to the linguistic context’. Reference interpretation is also driven by non-linguistic factors like pointing and gazing (Goodrich Smith & Hudson Kam, 2015; Nappa & Arnold, 2014), but written language relies more heavily on the linguistic context. Our study will not test between different possible mechanisms by which exposure might matter. Instead, we ask a more basic question: Is pronoun comprehension predicted by individual variation in print exposure?

Young children do not interpret pronouns like adults do

A second reason to suspect that exposure matters comes from research on pronoun comprehension in young children, which shows that they do not follow a subject-assignment strategy for pronouns as reliably or quickly as adults do. In Arnold *et al.*'s (2007) Experiment 1, three-and-a-half to five-year-old children listened to sentences like *Panda Bear is having lunch with Puppy. He wants a pepperoni slice*, and the experimenter asked, “Who wants the pepperoni slice?” Children responded by giving a toy pizza to one puppet, which revealed their interpretation of the pronoun. In this example, the two puppets are both male; other stories included one boy and one girl puppet. In the mixed-gender stories, children interpreted the unambiguous pronouns correctly. However, children in both age groups (3.5–4; 4–5) were completely at chance in the ambiguous pronoun condition, choosing both subject and non-subject equally. Koster, Hoeks, and Hendriks (2011) report a similar lack of a subject bias for children aged four to six, using a task with longer stories.

Similar results emerge from eye-tracking studies. In a study with four- to five-year-olds (Arnold *et al.*, 2007, Experiment 2), children viewed a picture of two characters and listened to a story, e.g., *Donald is bringing some mail to Mickey ... He is holding an umbrella ...*. The context either had two same-gender characters (making the pronoun ambiguous), or two different-gender characters. The target phrase was initially ambiguous, e.g., both characters were holding something, but only one held an umbrella. The authors examined fixations prior to the disambiguating word (*umbrella*) for evidence of a bias to look at the target word. In the different-gender conditions, children looked at the target as rapidly as adults did in Arnold *et al.*

(2000). But in the same-gender conditions there was no bias to look at the subject. This suggests that children do not follow the same subject bias as adults do, or at least not quickly enough to affect their eye-movements in this task. This finding is not limited to four- to five-year-olds, and in fact a similar pattern emerged for children aged eight to twelve (Arnold, Nadig, Bennetto, & Diehl, 2009).

The above studies show virtually no use of the subject bias by young children. By contrast, other studies have found evidence for a subject bias in very young children, but that they are much slower than adults to use it. For example, Song and Fisher (2005, 2007; see also Pyykkönen, Matthews, & Järvikivi, 2010) found that very young children (age 2.5 or 3) tended to look at the subject character following a pronoun in a story about two characters. Nevertheless, children in that study took a long time to exhibit a subject bias – not until 1–2 seconds after the pronoun – while adults in other studies looked rapidly, around 400 ms after pronoun onset (e.g., Arnold *et al.*, 2000). This led Hartshorne *et al.* (2015b) to propose that young children may know that subjects are good referents for pronouns, but are slower to apply this information during online processing. Hartshorne *et al.* tested this in an eye-tracking task similar to the one used by Arnold *et al.* (2007), except that the pronoun was fully ambiguous in their stimuli (e.g., *Emily went to school with Hannah. She read ten books.*) Five-year-olds looked at the subject character about 1400–1500 ms after the pronoun, while adults in their study took only 1100–1200 ms. They concluded that children differ from adults only in processing speed and not general strategy.

In a different approach, van Rij *et al.* (2013) implemented a computational model of pronoun comprehension. This model took the previous subject as the ‘topic’ of the discourse, and assigned pronouns to the topic. This effect was modulated by an implementation of working memory, which affected the activation of the subject representation. They proposed that their model accounted for Koster *et al.*’s (2011) findings that children were less likely to select the previous subject than adults, under the assumption that children have lower working memory than adults. That is, this approach is similar to Hartshorne *et al.*’s (2015b) in that it suggests that children and adults have the same model for pronoun interpretation, and differ only in cognitive capacity.

In sum, all previous work suggests that children aged two to five interpret pronouns differently from adults. Some studies find no evidence of a subject bias (Arnold *et al.*, 2007), whereas others find that children apply the subject bias more slowly than adults typically do (Hartshorne *et al.*, 2015b; Song & Fisher, 2005, 2007). The contrast between children and adults could be explained by the exposure hypothesis, which suggests that young children have not solidified their ability to access and use the subject bias for understanding pronouns, perhaps because it takes time to build strong representations of probabilistic biases and practice applying them. An alternative explanation builds on the assumption that children have a slower processing capacity (Hartshorne *et al.*, 2015b) or lower working memory (van Rij *et al.*, 2013).

However, it is important to note that these exposure and processing explanations are not mutually exclusive. It seems quite likely that children differ from adults in processing speed. Yet this does not rule out the possibility that exposure may also guide the development of discourse comprehension mechanisms. In the current study, we do not seek to contrast these explanations, but instead test the exposure hypothesis directly by examining individual differences in print exposure.

Print exposure affects pronoun comprehension for adults

A third reason to expect that exposure matters for pronoun comprehension comes from research with adults, which shows that adults with greater print exposure are more likely to follow the subject bias than adults with less print exposure. Arnold, Strangmann, Hwang, Zerkle, and Nappa (2018a) tested how individual differences in print exposure predicted variation in pronoun comprehension. In Experiments 1 and 2, they used a spoken pronoun comprehension task developed by Nappa and Arnold (2014; based on Arnold *et al.*, 2007, Experiment 1), where listeners could use either the linguistic context (i.e., the subject bias) or the speaker's gaze as a cue to the pronoun's referent. In this task, participants watch videos of a woman telling stories about four characters: Panda Bear (male), Puppy (male), Bunny, (female), and Froggy (female) (see Figure 1). Each story mentions two characters, e.g., *Panda Bear is having lunch with Puppy. He wants a pepperoni slice.* On the table are puppets for Panda Bear, Puppy, and a toy to represent the object in the story (here, the pizza). Each story appeared in three conditions, rotated across lists. In the neutral condition, at pronoun onset the speaker gazes at a toy pizza, which is uninformative about the pronoun referent. In the gaze-to-subject or gaze-to-non-subject conditions, the speaker gazes at one puppet or another as she says the pronoun. The gazes are not subtle, and involve a shift of the head, shoulders, and upper body. Thus, the question is not whether people notice the gaze gestures (they are hard to miss), but rather how they use them to interpret the pronoun. Participants respond to a question like, "Who wants a pepperoni slice?", which shows their interpretation of the pronoun. These responses are off-line, and subjects can take as long as they need to answer. Thus, processing speed should not affect responses in this task.

Arnold *et al.* (2018a, Experiment 1) found that subjects followed a subject bias in the neutral condition, selecting the subject character 86% of the time. This bias increased to 93% when the speaker gazed at the subject character (here, Panda Bear), and dropped to 67% when the speaker gazed at the non-subject character (here, Puppy; see also Nappa & Arnold, 2014, for similar gaze effects). Critically, they also measured individual differences in exposure to print materials, using the Author Recognition Task (aka the 'ART'; Acheson, Wells, & MacDonald, 2008; Moore & Gordon, 2015; Stanovich & West, 1989). The ART is a proxy measure for the amount that people read, and is measured by the number of authors people know from a list of real and non-real author names. Participants with higher ART scores had a higher rate of selecting the subject character, across all three conditions, than participants with lower ART scores. By contrast, there were no effects of individual differences in working memory (the automated operation span task; Unsworth, Heitz, Schrock, & Engle, 2005) or theory of mind (the mind in the eyes task; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). Experiments 2 and 3 found similar print exposure effects, but no meaningful effects of socioeconomic status.

Even though Nappa and Arnold (2014) had also tested pointing conditions, Arnold *et al.* (2018a) restricted their analysis to the gaze conditions in the critical items, although all fillers included pointing. Their motivation was that the gaze conditions elicited variation across individuals, and thus provided a good testing ground for individual differences. By contrast, the pointing conditions provided a very strong cue to pronoun interpretation, and less individual variation. Nevertheless, the effect of print exposure is not limited to experiments that include gazing manipulations: Arnold *et al.*'s (2018a) Experiment 3 included only auditory stories, and still exhibited an effect of print exposure.



Figure 1. Illustration of the pronoun video task (Nappa & Arnold, 2014) showing the three different conditions. Panel A depicts the gaze-to-subject condition; Panel B shows the neutral gaze condition, where the speaker gazes at the pizza; Panel C indicates the gaze-to-non-subject condition. (A version of this image was originally published in Nappa & Arnold, 2014.)

These findings demonstrated that individual differences in language exposure were correlated with individual differences in how people used the linguistic context to understand pronouns. Critically, this showed that print exposure affects spoken language comprehension, which goes beyond previous evidence that frequent reading improves written language comprehension (e.g., James, Fraundorf, Lee, & Watson, 2018).

Language exposure affects the development of vocabulary and grammar

A fourth reason to expect that exposure might affect pronoun comprehension is that we know that language exposure affects other aspects of language development. However, to our knowledge none of this work addresses spoken pronoun comprehension. The majority of evidence for exposure effects comes from measures of vocabulary knowledge or syntactic development, rather than higher-level discourse processes. Thus, our study fills a gap in the literature.

For example, several studies have found that print exposure predicts syntactic knowledge or processing. Montag and MacDonald (2015) found that children and adults with greater print exposure were more likely to produce passive relative clauses, a complex syntactic structure. They also demonstrated with a corpus analysis that this structure occurs more often in written than spoken language. There is also variation in how well adults process syntactic structures as a function of their print exposure (e.g., Farmer, Fine, Misyak, & Christiansen, 2017; James *et al.*, 2018). In addition, exposure within the context of the experiment itself facilitates the processing of rare syntactic constructions for adults (Fine & Jaeger, 2013; Wells, Christiansen, Race, Acheson, & MacDonald, 2009).

Another line of research suggests that language development in children is influenced by the complexity of language heard in the home. For example, vocabulary at 54 months is correlated with the quantity of speech produced in samples between 14 and 18 months (Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina, & Trueswell, 2013; see also Huttenlocher, Haight, Byrk, Seltzer, & Lyons, 1991; Rowe, 2008, 2012). Syntactic development is also correlated with the syntactic complexity of parental speech (Huttenlocher, Vasilevya, Cymerman, & Levine, 2002).

The quality and quantity of maternal speech has been argued to account for the effects of socio-economic status (SES) on language development. Several studies have found that SES correlates with language skill (see Hoff, 2013, for a review). It is well established that SES correlates with vocabulary knowledge (e.g., Dollaghan *et al.* 1999;

Hart & Risley, 1995; Hoff-Ginsberg, 1998) and grammatical development (e.g., Dollaghan *et al.*, 1999; Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Morisset, Barnard, Greenberg, Booth, & Spieker, 1990). The dominant explanation for these correlations rests on the role of maternal language. The idea is that mothers from low-SES families tend to produce less language overall and less complex language than mothers from high-SES families, and it is this input that explains variation in language development, and not SES itself. In support of this idea, evidence shows that socioeconomic status correlates with the complexity of child-directed speech at home (Fernald & Marchman, 2011; Hart & Risely, 1995; Hoff, 2003; Huttenlocher *et al.*, 2002), and in some cases maternal speech mediates the effect of SES on language development (Hoff, 2003). For example, in a longitudinal study, Rowe (2012) found that the quantity and quality of vocabulary in parental speech predicted vocabulary one year later, controlling for SES and other factors. Thus, this line of work also supports the idea that language exposure is important for language development.

The majority of work on SES and language development has focused on knowledge of vocabulary and grammar. Nevertheless, there is reason to think that SES may affect narrative development too, but findings at the discourse level are more limited. It has been shown that low-SES children produce less sophisticated narratives than higher-SES children (Heath, 1983; Vernon-Feagans, Hammer, Miccio, & Manlove, 2001), less informative referential expressions, and are less likely to identify ambiguous descriptions (Lloyd, Mann, & Peers, 1998), and are less likely to produce ‘contingent discourse’, which means the production of utterances that are related to each other (Hoff-Ginsberg, 1998). Most of these findings focus on broad measures of discourse production, with no evidence about pronoun comprehension. Nevertheless, such a link is possible. Given the known association between SES and language input, this could predict that children from high-SES families also hear more complex narrative structures than those from low-SES families.

The relation between SES and spoken language input raises questions about our current goal to test how print exposure relates to individual differences in pronoun comprehension, given that we did not collect measures of either SES or spoken language experience. We assume that language development is guided by all types of language exposure, both spoken and written. If children with high print exposure tend to come from high-SES families, and if high-SES families also produce more complex spoken discourses, we can’t know whether our individual differences stem from print exposure per se, or exposure to spoken discourse, or a combination of both. Thus, our goal is broader. We test print exposure as one indicator of individual differences in language exposure, acknowledging that it may be correlated with other types of language exposure. Our approach has the practical advantage that we can easily obtain a proxy for print exposure through the Title Recognition Task (TRT). In addition, we asked parents for estimates of children’s reading behaviors as an additional estimate of print exposure. By contrast, it would be extremely costly to measure individual differences in spoken language experiences.

Do children use social cues to guide pronoun interpretation?

The current study uses the gaze videos from Nappa and Arnold (2014) to test whether children also show evidence of print exposure. This allows us to ask a secondary question, namely how gaze affects pronoun comprehension in children. Relatively

little is known about this question. Yow (2013) used a similar task (which, like the Nappa & Arnold task, was modeled after Arnold *et al.*, 2007), and demonstrated that deictic pointing gestures guided pronoun interpretation for four-year-olds. These findings are consistent with evidence that very young children can use both deictic and iconic gestures for other communicative purposes (Behne, Liszkowski, Carpenter, & Tomasello, 2012; Goodrich & Hudson Kam, 2009; Meyer & Baldwin, 2013; Morford & Goldin-Meadow, 1992; Namy, Campbell, & Tomasello, 2004; Stanfield, Williamson, & Özçalışkan, 2014). On the other hand, Goodrich Smith and Hudson Kam (2015) tested whether children could learn that a gesture to a point in space was representative of a referent, and use this kind of gesture for pronoun comprehension. They found that older children (7–8) used both gestures and the subject bias, but younger children (4–5) did not use either cue.

By contrast, there is little work examining how children use gazing in the context of complex discourses. We do know that very young children (aged 1–2) learn to follow and interpret the gaze of another person (Deak, Krasno, Triesch, Lewis, & Sepeta, 2014; Woodward, 2003; Yu & Smith, 2013), that they can use this for learning new words (Baldwin, 1991), and that gazing has a referential component (Behne, Carpenter, & Tomasello, 2005; Csibra & Volein, 2008; Moll & Tomasello, 2004). Okumura, Kanakogi, Kanda, Ishiguro, and Itakura (2013) suggest that gaze shifts are taken by infants as “nonverbal deictic signals”. However, these experiments typically present babies with simplified contexts, often with only one object in the scene, even if it was sometimes hidden from view (e.g., Csibra & Volein, 2008). Infants can use gaze to identify which of two objects the speaker is referring to (Baldwin, 1991, 1993). Yet even in this case, the speaker uses gaze to signal attention to one single object, and never considers the other.

By contrast, gaze presents a more complex cue when combined with spoken narratives. Speakers mention many things in narratives, not just one. Some of these things are not physically present, and the speaker may gaze at co-present objects that are unrelated to the content of the speech. For example, a mother may be gazing at her child’s lunch while she packs it, simultaneously asking “Did you remember your homework?” Even if eye-gaze is directed at a discourse-relevant object, gaze may occur prior to the reference, or after, or not at all. This is especially true when there are two referents in the same sentence, which makes it likely that the speaker could mention one person while gazing at the other. For example, a speaker may gaze at Bill while saying “John hit Bill when he got mad”. Thus, gaze is not a simple cue to identifying a referent in running speech. This raises questions about the age at which children begin to use gaze as a cue for interpreting pronouns, and how much children of different ages rely on gaze vs. linguistic cues.

The current study

The goal of the current study is to test whether print exposure in children is related to individual differences in pronoun comprehension, in a task where both subject-hood and gazing provide possible cues. Previous research suggests that four- to five-year-olds are not adept at using the subject bias (Arnold *et al.*, 2007), and may not consistently follow pointing cues either (Goodrich Smith & Hudson Kam, 2015). We therefore take this age as our starting point, and examine how responses change as a function of both chronological age and print exposure throughout the elementary school years. We focused particularly on ages five to seven, recruiting

23 children in this group. We also included smaller samples of children in age groups eight to nine ($n = 17$) and ten to fourteen ($n = 15$).

While Arnold *et al.* (2018a) tested print exposure with the ART, we used a children's version of this task, the Title Recognition Task (TRT; Montag & MacDonald, 2015). This task asks children to recognize real children's book titles (e.g., *The Giver*) out of a list that includes fake ones (e.g., *The Last Shoe*). Like the ART, this task offers a reliable proxy for reading exposure. It correlates with both reading skill (Cunningham & Stanovich, 1990) and other language skills, like the rate of passive constructions produced (Montag & MacDonald, 2015). In addition, it is short and easy to administer in a few minutes. The final score is calculated as the total # correct minus # false positives.

Experiment

Methods

Participants

A total of 65 children between the ages of 5;1 and 14;6 participated. We recruited participants from the Morehead Planetarium afterschool program at UNC Chapel Hill, from the Orange County Library in Hillsborough, NC, from the Durham County Library main branch in Durham, NC, and through Chapel Hill neighborhood listservs. Data from 10 participants were excluded from analysis due to technological problems ($n = 3$), the parent reported that the child had a reading disorder ($n = 5$), or the child failed to meet criterion on the critical filler comprehension questions ($n = 2$). The analysis included 55 children in three age groups (see Table 1). In exchange for participation, children were given a choice of a book from our collection of age-appropriate books. We did not limit multiple children from the same family from participating in the study.

Of the children included in the analysis, one was reported to have cochlear implants, but the parent reported that the child had normal language development and school performance. Three children were reported to have attention deficit disorder. No children were reported to have uncorrected visual deficits, memory disorders, or autism spectrum disorder.

Materials

Our study included two questionnaires for the parent, the pronoun video task, and the Title Recognition Task.

Background questionnaire

Parents filled out a voluntary background questionnaire with information about the child's date of birth, grade, sex, ethnicity, race, number of languages spoken and proficiency, and any disorders in hearing, language, vision, memory, or attention. We also asked for the child's Lexile score¹ from the last report card, but most parents did not fill this out, so this measure is not included in our analyses.

¹A standardized reading measure used in North Carolina schools.

Table 1. Demographic data for participants. In two cases we knew the age of the child in years but not months, and estimated the months at 6. We counted a child as bilingual if the parent reported that they knew a second language at all, even if only a little (these languages included Spanish, Korean, and Chinese). The percentage bilingual and Hispanic/Latino are calculated out of those participants who reported this information on the questionnaire.

	N	Age range	Avgage	# F	Bilingual % (# reported)	Hispanic / Latino % (#reported)	Ethnicity				
							White	Black	Asian	More than one	Did not report
5–7 y.o.	23	5;1 to 7;9	6;6	15	22% (20)	9% (19)	16	5	0	1	1
8–9 y.o.	17	8;2 to 9;11	8;11	10	41% (15)	18% (15)	12	1	1	0	3
10–14 y.o.	15	10;2 to 14;6	11;9	10	10% (10)	33% (11)	4	4	1	1	5
Total	55	5;1 to 14;6	8;8	35	35% (45)	18% (45)	32	10	2	2	9

Who wants a pepperoni slice?

Puppy

Panda Bear

Figure 2. Example question screen.

Reading questionnaire

Parents were asked to report the number of hours that the child heard books read aloud, read to him/herself, and listened to books on CDs, tapes, or other media. These estimates were requested for three time periods: (1) over the last three months; (2) when the child was five years old; and (3) when the child was three years old. We also asked whether the child enjoyed reading, or hearing books read aloud, both on a 1–4 scale (see ‘Appendix A’).

Pronoun video task

We used a subset of the video materials from Nappa and Arnold (2014, Experiment 1; example videos available at <<http://arnoldlab.web.unc.edu/publications/supporting-materials/nappa-arnold/>>). In each video, a woman sat at a table with two stuffed animal characters on either side, and a toy prop in the middle. Four characters appeared in the stories: Puppy (male), Panda Bear (male), Bunny (female), and Froggy (female). In all critical items, the two characters were the same gender. Each story began by reminding the child of the character’s names: “This story is about X and Y. This is X, and this is Y.” The speaker then paused and put her hands to her lips, indicating that the story was about to begin. For example, in Figure 1 the story scene depicts Puppy, Panda Bear, and a toy pizza. The story consisted of two sentences; for Figure 1 the story was “Panda Bear is having pizza with Puppy. He wants a pepperoni slice”. When the video ended, a new screen presented a written question, such as “Who wants a pepperoni slice?” The two options were presented vertically (see Figure 2). The order of presentation was counterbalanced across items, such that the top answer was correct on 40% of the fillers, and was the subject character name on 8 out of the 15 experimental stimuli.

For the majority of the subjects, the pronoun task included 5 stimuli in each condition, and 10 fillers, for a total of 25 (plus 2 practice). The first two participants (ages 7 and 9) saw a longer version of the experiment, with 7 stimuli in each condition, and 21 fillers, for a total of 42. Although they did manage to pay attention for the whole task, it was clear that the length might be challenging for some children, so we reduced the length for subsequent participants.

We only used the gazing (not pointing) conditions from Nappa and Arnold (2014), so the videos occurred in three conditions, as shown in Figure 1. The gaze manipulation was more than just eye-movement, and also included shifts in the head and shoulders. Thus, our gaze manipulation was visually salient and hard to miss.

The spoken story was identical across conditions, but the speaker gazed either to the subject character, the non-subject character, or the neutral object (the pizza) in the center. The filler items took a similar form, but never used pronouns and always used a felicitous point, which means that the reference was fully disambiguated. Additionally, seven of the ten fillers had two characters of different gender. Four of the fillers used linguistic structures that were identical to the experimental stimuli, except they used a repeated name in the last sentence, and six of the fillers

introduced the characters in a coordinate NP structure like “Panda Bear and Puppy are getting ready for school. Panda Bear wants to wear the jacket”. Three of the fillers asked a question about the story, e.g., “What does Panda Bear want to wear?” (the jacket / the shirt). Seven of the fillers asked about the location of one character, e.g., “Who was on the left side of the screen?” Each of the four characters appeared as the grammatical subject of the critical items three or four times, and as the first-mentioned character of the fillers two or three times.

Title Recognition Task

We used the Title Recognition Task from Montag and MacDonald (2015). Their version of the task was modified from the one used by Cunningham and Stanovich (1991), using books that are more recently popular. There were 30 real titles and 20 foils. The actual book titles on the list are all fiction books, and include picture-books appropriate for very young children (e.g., *Stellaluna*), easy-read books (e.g., *Poppleton in Winter*), as well as books typically read in a middle or high school English class (*To Kill a Mockingbird*; *Animal Farm*). The list was presented to the child on a piece of paper. If the child needed help reading it, the experimenter read it to them. Children were told that this was a quiz to see how many book titles they recognized. They were cautioned not to guess, and told that they would lose points for picking fake titles. Children made a check or X on a line next to each title to select it. The TRT score was calculated as the total number of correct selections minus the total number of incorrect selections.

Procedure

Parents were informed about the nature of the tasks, and asked to sign the consent form and fill out the background reading questionnaires. The experimenter greeted the child and explained that they would be completing a game listening to a woman tell very short stories, and then answer questions about books. Children aged seven and over were asked to read and sign an assent form; the experimenter read the form if needed.

For the pronoun video task, participants were told: “Today you’re going to watch some stories on videos and answer some questions about them. These stories were made for younger children, but we also want to know what older children think about them.” This explanation was necessary to explain the childlike nature of the videos. Even though our subjects were children, the videos are simple enough to be appropriate for preschool children, and consultation with children in this age range made it clear that the videos seemed ‘babyish’. We also told children that the questions were hard, so they had to pay attention. Children were told that the videos would be about four characters, Puppy, Panda Bear, Froggy, and Bunny, and that Puppy and Panda Bear are boys, while Froggy and Bunny are girls. Because some of the filler questions asked about left/right position of characters, the experimenter asked if the child knew right from left, and if the child was uncertain, the experimenter put post-it notes with the letters L and R next to the computer as an aide.

Children then saw a practice video with two stories. The first was similar to the critical items, but did not use a pronoun (*Puppy and Panda Bear are going to take some pictures together. Puppy wants to hold the camera first.*). Children saw a screen with the question “Who wants to hold the camera first?” with two options (see Figure 2). They were instructed to press the i button for the top choice, and the m

button for the bottom choice. These keys were covered with colored squares on the laptop keyboard to help children remember which keys to press. In both the practice and main task, the experimenter read the question aloud if the child needed help. The second practice video was followed by the question: “Who was on the left side of the screen?” The experimenter pointed out that some of the questions would be like that, so they should pay attention to the location of each character.

Children were then told that we were going to see if they remembered who was a boy and who was a girl. The experimenter started an E-prime task, which showed pictures of each character, followed by a screen with the options ‘boy’ and ‘girl’. For example, for the picture of Panda Bear the experimenter said: “This is Panda Bear. Is he a boy or a girl?” The child pressed the top option for boy, and the bottom option for girl. The experimenter asked if the child had any questions, and then started the main pronoun task.

Following the pronoun task, the experimenter said: “Now we are going to do a quiz to see what books you know.” The child was given the Title Recognition Task (Montag & MacDonald, 2015). They were told that they would lose points for wrong titles, so they should not guess. After they finished the TRT, the experimenter thanked them and invited them to pick out a book from our collection.

Results

We excluded trials if the research assistant noted a problem during the trial, for example if the video froze or if there was an interruption during the video (e.g., if the child spoke). A total of 11 trials were excluded, with no more than one per participant. For all analyses, our dependent measure was the likelihood of selecting the character in subject position as the response to the question for critical items, which reflected the child’s interpretation of the pronoun. For each analysis, we fit a multilevel logistic regression model with centered predictors and maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), using SAS proc glimmix with a binary distribution and a logit link. In order to distinguish between our three conditions: gaze-to-subject, gaze-to-neutral, and gaze-to-prepositional object, we used two dummy variables (gaze-to-subject: (1 for gaze-to-subject condition and 0 otherwise), and gaze-to-non-subject: (1 for gaze-to-non-subject condition and 0 otherwise)). All models included crossed random effects for both participant and item and random slopes for critical predictors with respect to both participant and item. When the model estimated a random effect to be zero, it was removed from the model. All predictors were centered in all models.

We examined our data using three complementary analyses: (1) a simple model of group results as a function of the gaze predictor; (2) an analysis including age and TRT score as predictors; and (3) a model in which predictors were factors extracted from an exploratory factor analysis of the responses to the parental questionnaire, specifically, those related to reading exposure.

Group results

We initially examined our data for our group of subjects as a whole. We found that children, like adults, were more likely to choose the subject character overall, and this effect was modulated by gaze: 85% (SE = 2%) in the gaze-to-subject condition, 70% (SE = 4%) in the neutral condition, and 61% (SE = 5%) in the gaze-to-non-subject

condition.² In a model with only the two gaze predictors and the control predictor of trial order, we found a significant effect of gaze-to-subject ($b = 0.99$ (SE = 0.27), $t = 3.61$, $p = .001$), a marginal effect of gaze-to-non-subject, $b = -0.41$ (SE = 0.23), $t = -1.77$, $p = 0.08$, and no effect of trial order ($b = 0.02$ (SE = 0.01), $t = 1.56$, $p = .12$).

Age and TRT score effects

Our focal question was whether these effects were modulated by our two key predictors, age and print exposure. As expected, these variables were correlated with each other, such that older children tended to have higher TRTs ($r = 0.61$, $p < .001$) compared to younger children; see Figure 3. TRT scores ranged from -4 to 20 , with an average of 5.6 . Figures 4 and 5 reveal that the rate of selecting the subject character was higher both for older children and for children with higher TRT scores.

We therefore began by examining age and TRT in separate models, and then combining the two in a single model. In each of the separate models with age and TRT, we first built a model with all main effects and interactions between the gaze manipulations and either age or TRT, and then removed any interactions that fell below the significance cutoff of $|t| = 1.5$. The results of these models are reported. We also included trial order as a control predictor, given evidence that biases may change over the course of an experiment (Fine & Jaeger, 2013), although it was not significant in any of our models. Table 2 reports the final models. In the combined model, we combined the predictors included in the two separate models.

In the model with only age, we found that age had a marginally significant positive effect on responses, such that older children were somewhat more likely to choose the subject response. Age did not interact with gaze manipulations. By contrast, in the model with only TRT, we found that TRT had a significant positive effect, such that children with higher TRT scores were more likely to choose the subject response. This also interacted with the gaze-to-subject manipulation. When both age and TRT were entered in the same model, only TRT and the gaze-to-subject*TRT interaction were significant, and the marginal age effect disappeared. Figure 3 suggests that the gaze-to-subject predictor had a stronger effect on children with lower print exposure than those with higher print exposure.

We probed the interaction between TRT and gaze-to-subject by computing conditional moderating variables (Holmbeck, 2002) and conducting two post-hoc multilevel logistic regression models with maximal random effects. One model enabled us to test the simple slope of the gaze-to-subject condition, whereas the other tested the simple slope of the neutral/gaze-to-non-subject conditions³ (see Figure 6 for simple slopes). These models were otherwise identical to model C, including predictors age, TRT, and trial order. We found that TRT is significant

²These means represent the average of participant means by condition, and the standard errors are calculated from these means in each condition.

³In the model testing gaze-to-subject condition, we used a binary predictor where 0 represented the gaze-to-subject condition, and -1 represented the gaze-to-other condition. This means that the intercept represents the average subject response in the gaze-to-subject condition, when the other centered predictors are zero. In the model testing the gaze-to-other, 0 represented gaze-to-other, and 1 represented gaze-to-subject, such that the intercept in this model represents the average subject response for the gaze-to-other condition. The coefficient of the centered Title Recognition Task predictor in each of these models is the simple slope of the regression line representing the relation between choosing the subject response and print exposure for each condition.

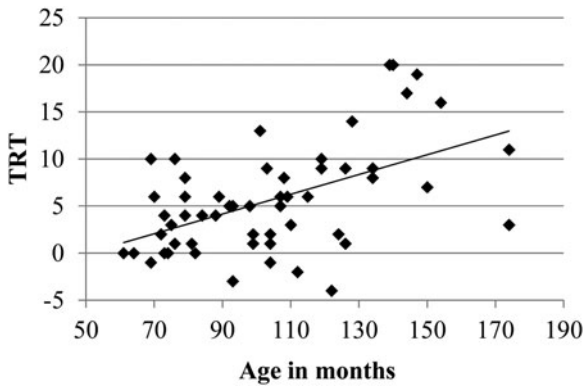


Figure 3. Correlation between age and score on the Title Recognition Task in our sample.

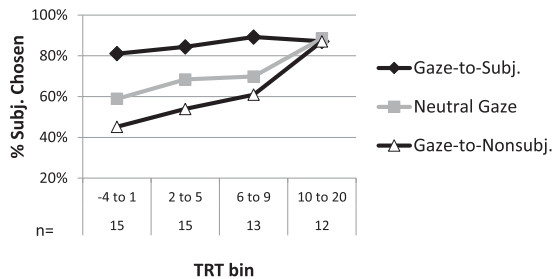


Figure 4. Experiment results, plotted as the average percentage of responses selecting the subject character, divided by TRT scores.

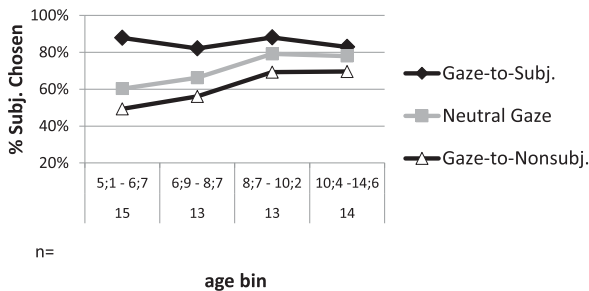


Figure 5. Experiment results, plotted as the average percentage of responses selecting the subject character, divided by age.

when gaze is neutral or to the non-subject ($b = 0.11$ (0.03), $t = 3.24$, $p = .002$), but not in the gaze-to-subject condition ($b = 0.02$ (0.04), $t = 0.38$, $p = .70$). These results suggest that print exposure effects may be weaker in the gaze-to-subject condition, when responses are generally high. We return to this issue in the ‘General discussion’.

In order to assess the unique contribution of TRT, we re-did model C and requested Type I sums of squares’ tests. In these tests, fixed effect predictors are added one by one

Table 2. Results of statistical models.

A. Model with Age				
<i>Effect</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	0.69	0.25	2.79	.006
Gaze-to-subject	1.0	0.28	3.59	.001
Gaze-to-non-subject	-0.42	0.23	-1.80	.076
Age	0.01	0.01	1.81	.076
Trial order	0.02	0.01	1.57	.12
B. Model with TRT				
<i>Effect</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	0.74	0.24	3.03	.003
Gaze-to-subject	0.90	0.28	3.28	.003
Gaze-to-non-subject	-0.45	0.23	-1.97	.052
Title recognition task (TRT)	0.11	0.03	3.78	.0003
Gaze-to-subject*TRT	-0.09	0.04	-2.21	.030
Trial order	0.02	0.01	1.52	.129
C. Combined model				
<i>Effect</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	0.74	0.25	3.02	.003
Gaze-to-subject	0.90	0.28	3.28	.003
Gaze-to-non-subject	-0.45	0.23	-1.97	.052
Age	0.002	0.006	0.29	.776
Title recognition task (TRT)	0.11	0.03	3.19	.002
Gaze-to-subject*TRT	-0.09	0.04	-2.21	.030
Trial order	0.02	0.01	1.52	.130

Note. In all three models, there was a random slope for gaze-to-subject and gaze-to-non-subject by subjects, but only a slope for gaze-to-subject by items; the gaze-to-non-subject slope by items was estimated to be zero and removed from the model.

and 1-degree of freedom *F*-tests are conducted to assess the contribution of each effect to the model. We specified the TRT and gaze-to-subject*TRT interaction as the last terms in the equation (in that order) so their *F*-tests would reflect the contribution above and beyond every other term in the model. Both TRT and its interaction with gaze-to-subject had a significant contribution to the model ($F(1,37.25) = 5.58, p = .02$ for TRT and $F(1,92.47) = 5.02, p = .03$ for the interaction term).

Exploratory factor analysis

Our final analysis sought to obtain an even better measure of print exposure by incorporating responses to the parental questionnaire. However, our first

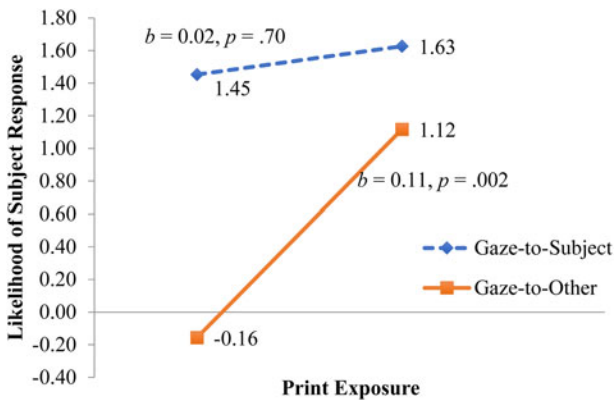


Figure 6. Simple slopes to illustrate the interaction between print exposure (TRT) and the gaze-to-subject manipulation.

consideration was to make sure that we analyzed questionnaire items that were behaving consistently across ages in our sample. Items that do not behave consistently are said to exhibit differential item functioning in the psychometrics literature and are thus dropped from analyses (de Ayala, 2013). To ensure consistency, we examined correlations amongst the parent questionnaire items and TRT (see Figure S1 in 'Appendix B'), and selected those items that exhibited a pattern of positive associations among themselves across all ages in our sample. We selected this approach because our sample is too small to enable estimation of item response theory models. Five items were selected as the focus of the exploratory factor analysis (EFA): the TRT scores (our objective measure of print exposure), in addition to parental reports of children's self-reading at three and five years of age and parental reports of how much children were read to when they were three and five years old. We included TRT in our EFA to triangulate across objective and subjective measures of print exposure and attain a more valid assessment than TRT by itself. This analysis only includes the 44 children whose parents completed all the questions selected for this analysis.

Using the principal axis extraction method with Promax rotation (Fabrigar, Wegener, MacCallum, & Strahan, 1999), we fit one- and two-factor models. We found support for two factors, which we labeled Print Exposure and Parent-led Early Reading. Self-reading at three and five years of age loaded together with the TRT score as indicators of the Print Exposure factor (standardized loadings = .65, .60, and .41, respectively), whereas being read to at three and five years loaded separately as indicators of Parent-led Early Reading (standardized loadings = .88 and .94, respectively). The two factors were positively correlated ($r = 0.66$). We estimated scores for each factor to include as predictors of subject-choice in the multilevel logistic regression model.

Our models included age and the two factors, Print Exposure and Parent-led Early Reading, as well as trial order. We also included interactions between gaze and the three individual difference predictors, but then trimmed the model by excluding interactions where $|t| < 1.5$. Given the scale and centering of the predictors, the intercept in our model represents the average probability of choosing the character in the subject position in the gaze-to-neutral condition for children of average age within our

Table 3. Estimates from multilevel logistic regression model for the prediction of subject choice

Effect	<i>b</i>	<i>SE</i>	<i>T</i>	<i>p</i>
Intercept	0.85	0.27	3.11	.002
Gaze-to-subject	0.75	0.34	2.21	.036
Gaze-to-non-subject	-0.47	0.28	-1.66	.101
Trial order	0.023	0.014	1.61	.108
Age	0.017	0.006	2.59	.013
Print exposure	0.85	0.33	2.55	.014
Parent-led early reading	-0.19	0.29	-0.65	.521
Gaze-to-subject*Print exposure	-0.78	0.33	-2.39	.020

sample (8;8) and who had average levels of Print Exposure and Parent-led Early Reading.

Results for the model are included in Table 3. As expected, the predicted probability for choosing the subject in the story for the average-aged child in the gaze-to-neutral condition with average Print Exposure and Parent-led Early Reading scores is positive and significant ($b = 0.88$, odds = 2.41, $p = .002$). The subject bias for this group increases within the gaze-to-subject condition ($b = 0.68$, odds = 1.98, $p = .054$), such that children are nearly twice as likely to pick the subject if they are in the gaze-to-subject condition when compared to the gaze-to-neutral condition. Moreover, there was a small but positive effect of age ($b = 0.02$, odds = 1.02, $p = .025$), suggesting a slightly higher rate of choosing the subject in the story for older children. The Print Exposure factor had a significant positive average effect too ($b = 1.00$, odds = 2.73, $p = .017$), pointing to a much higher tendency of picking the subject for children with higher-than-average scores in Print Exposure.

Finally, there was a significant interaction of gaze-to-subject condition * Print Exposure ($b = -0.78$, odds = 0.46, $p = .02$). As with the interaction in model C above, the interpretation of this interaction is for children of average age and average level of Parent-led Early Reading. Again we probed this interaction by testing the simple slopes of the regression lines representing relations between choosing the subject response and print exposure for the gaze-to-subject (Model 1) and the gaze-to-other (Model 2) conditions. For the gaze-to-subject condition, Print Exposure was not significant ($b = 0.04$ (0.37), $t = 0.11$, $p = .91$). This reflects the tendency for most participants to select the subject character in this condition. For the gaze-to-other condition, there was a significant positive effect of Print Exposure ($b = 0.79$ (0.33), $t = 2.37$, $p = .02$).

Given our focal interest in the print exposure measure, we again relied on the Type I SS tests from SAS for assessing the unique contribution of Print Exposure and its interaction with gaze-to-subject. Here, the last three terms we entered into the equation were Print Exposure, gaze-to-subject X Print exposure, and gaze-to-non-subject X Print exposure, in this specific order. The average effect of Print Exposure was not significant ($F(1,40.13) = 3.57$, $p = .066$), but its interaction with gaze-to-subject was ($F(1,66.54) = 4.82$, $p = .03$). As in the original model, the interaction with gaze-to-non-subject was not significant ($F(1,78) = 0.01$, $p = .91$). These results support

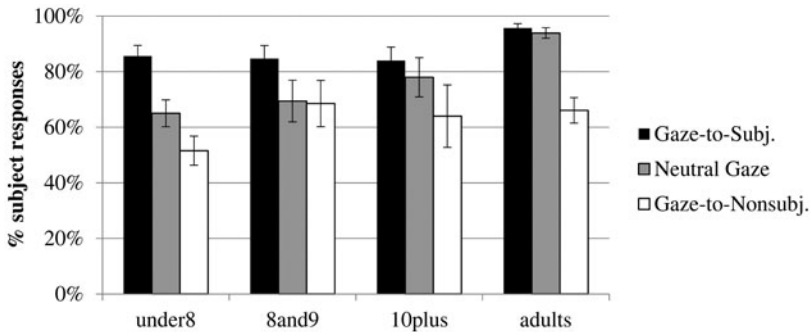


Figure 7. Average subject responses by age group in the current study, compared to adults (Arnold *et al.*, 2018a, Experiment 2). Reported means are averages of participant means by condition; standard error bars are calculated across participant means by condition.

once again the unique contribution of Print Exposure to the overall model while controlling for age and all other terms in the equation.

In this model, in contrast to the models with age and TRT in the previous section, we did find a significant effect of age. The difference may be due to the slightly different sample of children, as this model only included those children whose parents completed all the selected questions on the questionnaire, or it may be due to the influence of other factors in this model. In combination, these models suggest that there may be an effect of age, but it is not as consistently apparent as the effects of print exposure.

No additional effects in the model were significant, suggesting that average-aged children in the gaze-to-neutral condition did not differ from those in the gaze-to-non-subject condition with regard to their rate of choosing the character in the subject position of the story. Similarly, there was no significant effect of Parent-led Early Reading.

At what age does the subject bias emerge?

The current paper was not designed to identify the age when children begin to follow a subject-assignment strategy. Indeed, our primary finding was that print exposure guides the development of the subject-assignment strategy, and exerts a stronger effect than age. Because this is not the focus of our paper, we do not consider age effects statistically. Nevertheless, the debate in the literature raises questions about how our study compares with other reports about children's performance at different ages. To this end, we consider the numerical trends in our task across ages, and how this compares with adult data from the same task (Arnold *et al.*, 2018a, Experiment 2). Figure 7 shows that the rate of selecting the subject increases over development, particularly in the neutral condition, which reflects the linguistic bias most clearly.

Notably, the effects of print exposure seen in this study are very similar to those measured for adults by Arnold *et al.* (2018a) using the same task. One difference is that, for children, the TRT measure interacted with the gaze-to-subject manipulation, but for adults, their measure of print exposure (the ART) did not interact with gaze. However, the gist of the results is the same – both children and adults use the subject bias more consistently if they have greater print exposure. Together, this set of experiments provides strong support for the exposure hypothesis.

In principle, language exposure could explain the difference between young children and adults. For example, Arnold *et al.* (2007) found that three- to five-year-olds showed no significant use of the subject bias at all, whereas adults as a group show a strong subject bias (e.g., Arnold *et al.*, 2000). If exposure explains everything, this effect could reflect the fact that very young children have had relatively little print exposure, due to their young age. As people age, the opportunity to learn from discourse exposure increases, such that most adults have acquired a robust use of the subject bias. Nevertheless, people at all ages exhibit variability in their use of the subject bias, and this variability corresponds to their print exposure. This finding does not rule out the possibility that processing speed changes over development, and that this may also explain some differences between children and adults (Hartshorne *et al.*, 2015b), but it does clearly support a role for exposure.

Analysis summary

In both of our approaches to examining print exposure, this construct was a robust predictor, such that children with more print exposure tended to select the subject character more than children with less print exposure. In our simple models, print exposure was indexed only by TRT scores. In the models with EFA factors, the print exposure factor included both TRT scores and parental estimates of self-reading at ages three and five. Print exposure also interacted with the gaze-to-subject manipulation, such that children with less print exposure showed a stronger gaze-to-subject effect than those with more print exposure. We also found that older children had a stronger bias than younger children to select the subject character, although this was only significant in the model with print exposure factors extracted from the exploratory factor analysis.

General discussion

Our experiment provided systematic evidence about how children aged five to fourteen interpret ambiguous pronouns. Our task carried two potential sources of information about the pronoun referent: (1) the linguistic context, where the subject of the previous sentence was the preferred referent of the pronoun; and (2) social gaze cues toward discourse referents. Like adults, children in our task generally tended to follow a subject-assignment strategy. Our focal question was whether the subject bias or gaze effects would vary as a function of reading exposure, controlling for expected age effects.

Our primary finding was that print exposure is correlated with the pronoun comprehension strategies used by young children. We used the Title Recognition Task as a proxy measure of exposure to books and found that it was a robust predictor of responses. Children with greater print exposure were more likely to pick the subject character as the referent of the ambiguous pronoun. This effect appeared in both the combined analysis of age and TRT effects, and also in the model using factor scores of print exposure. It is notable that our objective measure of print exposure, TRT, loaded together with parent reports of the child's reading habits at ages three and five in the exploratory factor analysis. Despite the fact that these reports may be noisy and subject to memory demands, parental reports patterned with children's current knowledge of titles. In addition, we found that the effect of TRT was modulated by an interaction between TRT and gaze-to-subject, where TRT

had less of an effect when the speaker gazed at the subject character. That is, when the speaker provided a social cue (gaze) that supported the subject bias, print exposure had less of an effect.

A second finding of interest was that age had a minor but less robust effect on responses. Numerically, older children tended to select the subject more than the younger children. In keeping with this, age was a significant positive predictor in the model with factor scores. However, age was not significant in the combined analysis with TRT, even though this sample was slightly larger than the model with factor scores. This is consistent with the conclusion that age does matter – as children grow older, they are more likely to converge on a regular subject-assignment strategy. However, the age effect was relatively small, and its effect was sometimes overshadowed by print exposure, which is correlated with age.

In addition, we found robust sensitivity to the gaze manipulation across all ages. There was a significant effect of the gaze-to-subject predictor. This shows that gaze to the subject increased selection of that character over the neutral condition. The effect of gaze-to-non-subject was smaller, and only approached significance in the combined Age/TRT model, and not the model with factor scores. Gaze did not interact with age, suggesting that the effect may be consistent across the elementary school years.

An open question in the literature is whether young children follow a subject-assignment strategy or not. Some studies fail to find this effect in four- to five-year-olds (Arnold *et al.*, 2007), but some studies have detected a subject bias in two- to five-year-olds, using tasks that allow plenty of processing time (Hartshorne *et al.*, 2015b; Song & Fisher, 2005). Our study examined an older age group, and found that most children exhibited a numerical subject bias. However, we also found significant effects of print exposure. This raises a question about the participants in earlier studies. If participants across experiments had different average levels of print exposure, this might explain differences in their findings.

One question about the observed print exposure effect is whether it stems from variation in children's use of the subject bias, or whether it instead reflects children's tendency to attend to social cues. For example, if children who read a lot are less likely to engage socially with peers, they may not develop a sensitivity to gaze gestures as an informative signal. Alternatively, it may be that all children follow the subject bias when there are no competing cues, but in the presence of competing gaze cues, perhaps children with less language exposure tend to favor the visual cue. Given the interaction between gaze-to-subject and TRT, both interpretations are possible. In fact, Figure 4 suggests that the highest-TRT group was relatively unaffected by gaze. However, there are two reasons to think that our finding reflects variation in sensitivity to the subject bias, rather than variation in sensitivity to gaze. First, responses in the neutral condition correlated with TRT ($r = 0.34$, $p = .01$). The neutral condition carried no information about gaze, so gaze cannot account for this effect. Second, research with adults shows that print exposure predicts responses even in a task without gaze cues (Arnold *et al.*, 2018a, Experiment 3). Further research is needed to establish whether print exposure itself affects sensitivity to gaze (which seems implausible), or whether decreased gaze sensitivity only occurs in the context of increased sensitivity to the subject bias.

Our findings suggest that exposure shapes the strategies that children use to interpret ambiguous pronouns. This raises questions about what kind of exposure matters. Is it reading itself that matters, or can listening to books have the same effect? Our

exploratory factor analysis (EFA) sheds some light on this question. We examined parental estimates of how much children read to themselves at ages three and five, and how much they were read to at these ages. As shown in 'Appendix B', these measures were highly correlated with each other. However, the EFA found that self-reading loaded onto the same factor as the TRT scores, but parent-led reading did not. This suggests that reading behavior early in life covaries with children's familiarity with books at older ages, which supports the assumption that the TRT is a proxy for print exposure, and that the primary method for exposure to books is through reading. A question for future research is whether pronoun comprehension is also influenced by variability in the type or complexity of print exposure.

It may seem surprising that parent-led reading (i.e., reading to kids) did not load onto the same factor as self-reading, and that it did not significantly predict subject responses. Reading skills are typically either absent or just emerging in three- and five-year-olds, so 'self-reading' does not represent the same thing as it does for adults. Nevertheless, it is well established that reading habits are supported by early reading behaviors. For a child, 'reading' might mean looking at a book, or talking about the pictures. These practices support a positive relationship with books, and help children develop reading skills. Thus, it is not surprising that early self-reading is related to the TRT measure of print exposure.

On the other hand, parent-led reading at ages three and five did not have a direct relationship with later measures of pronoun comprehension. Parent-led reading was positively correlated with self-reading ($r = 0.77$, $p < .001$). Thus, early parent-led reading may support later self-reading behaviors. Yet it was only the print exposure factor (which includes self-reading at ages three and five and the TRT) that predicted pronoun comprehension. The print exposure factor correlated with each participant's average subject response for both the neutral condition ($r = .42$, $p = .005$) and the gaze-to-non-subject condition ($r = .37$, $p = .01$), but the parent-led-reading factor did not correlate with responses in any condition ($r_s < 0.22$, $p_s > .16$). We speculate that self-reading is important because children can amass a greater amount of print exposure if they can read on their own, given that parent-led reading is limited by the parent's availability. By contrast, parent-led reading may not be frequent enough to strongly affect discourse comprehension development.

The current results strongly indicate that language experience, measured here with print exposure, is related to the cognitive mechanisms used to interpret pronouns. Yet this finding is fundamentally correlational, which necessarily raises questions about whether it can be explained by something else, such as socioeconomic status. It is well known that reading skill correlates with the availability of books and educational resources, which are dependent on economic resources (Bradley, Corwyn, McAadoo, & García Coll, 2001; Orr, 2003). This question is relevant for considering what kind of cognitive mechanism might explain our findings. If our findings are 'merely' a reflection of socioeconomic differences, it might be the case that some other cognitive correlate explains response differences. For example, low-socioeconomic status children may be less likely to pay attention during the task (Russell, Ford, Williams, & Russell, 2016). However, this explanation is unlikely, for two reasons. First, we excluded children who could not successfully answer our critical filler questions, which were unambiguous. This means that all children exhibited a certain level of attentiveness in our study.

Second, although we did not measure SES in this study, we expect that it may have been difficult to detect SES effects within a sample such as ours. Our children were

recruited from literate populations, including libraries and a university afterschool program. Within this population, SES differences may be slight. Moreover, when we tested for SES in our parallel adult study (Arnold *et al.*, 2018a), we found it did not explain pronoun comprehension biases.

Nevertheless, ultimately we cannot rule out the possibility that print exposure is correlated with other relevant individual differences. The most likely additional source of variability could be variation in spoken language input. Even beyond SES effects, there may be variability in the parental language that correlates with the amount of reading taking place in a family. But importantly, if such an effect were to exist, it would point to the same conclusion – namely, exposure matters. Thus, our findings make an empirical contribution by demonstrating a relationship between one measure of language exposure, print exposure, and the use of linguistic context for pronoun comprehension.

At the same time, we speculate that exposure to books may provide a particularly good domain for learning discourse patterns. We do not know precisely what types of exposure facilitate pronoun comprehension, but there are several likely possibilities (also see Arnold *et al.*, 2018a for discussion of this point). At a broad level, children may need exposure to highly cohesive examples of discourse in order to practice seeking within-discourse connections to understand how each utterance relates to the previous discourse. Printed language tends to be highly cohesive. Spoken language can also support this process, but sometimes spoken language is more related to the physical context (e.g., *Look at that!*) than to the previous discourse. Alternatively, children may need practice establishing anaphoric referential links – that is, recognizing that a referential expression co-refers with a previously mentioned one. Again, printed language is a good domain for this effort, because it tends to be internally cohesive.

Alternatively, as suggested by probabilistic theories of reference comprehension (Arnold, 1998; Frank & Goodman, 2012; Hartshorne *et al.*, 2015a; Kehler & Rohde, 2013), children may need to learn which patterns of reference are more frequent. For example, they may need to learn that referential expressions are likely to co-refer with the previous subject. This pattern has been demonstrated with analyses of children's books. Arnold *et al.* (2018b) examined five children's books, in a reanalysis of Arnold's (1998) text analysis. They examined all clauses that included more than one referential expression, and asked how frequently the subject was mentioned again in the following utterance vs. how often non-subject entities were mentioned. They found that subjects were more likely to be mentioned again than non-subjects. This reflects the view that the subject position is topical (Chafe, 1976). If children learn that subjects are likely to continue to be important in the discourse, they may develop an expectation for reference to the subject, which would facilitate subsequent comprehension.

Note that this generalization occurs for all expressions, not just pronouns. This means that exposure to all referring expressions, and not just pronouns, could potentially teach children about the salience of the subject position. However, this may be easier to learn in the context of complex utterances. All sentences contain a subject. Thus, by observing sequences like *The robbers ran, and they ran* (from *Georgie and the Robbers* by Robert Bright, 1963), children might conclude that speakers are likely to refer to recently- mentioned entities (which happens also to be true), and not specifically that they tend to refer to subjects. By contrast, more complex sequences illustrate that subjects are relatively more likely to be referred to than other entities, as in *As soon as Encyclopedia finished breakfast, he printed fifty*

handbills (from *Encyclopedia Brown*, by Donald J. Sobol, 1963). Here, readers can observe that the second sentence refers to the subject referent (*Encyclopedia*) and not the non-subject referent (*breakfast*). While complex sentences can occur in both spoken and written language, they are likely to be more common in writing. Nevertheless, future work is needed to tease apart the contribution of spoken and written language to the effect of print exposure on pronoun comprehension.

In sum, language exposure provides people with evidence about reference, and specifically that the subject entity is likely to get mentioned again. This pattern is probabilistic, so people may represent it in varying degrees of probability. Some people may in fact overgeneralize, generating a high expectation of subject reference, while others may not expect subject reference as strongly. Thus, one possibility is that individuals vary in the STRENGTH of the subject bias in their mind. In addition, individuals may also vary in the relative WEIGHT that they put on the linguistic context in general, and the subject bias in particular. Thus, in real discourse understanding, some people may follow social cues like gaze, or semantic cues, relatively more than a structural bias toward the subject. Based on our current results, both strength and weight explanations are possible.

While there are open questions about exactly what people learn from language exposure, this paper provides direct evidence that one type of exposure – reading – is correlated with pronoun comprehension strategies in children. Critically, our study tested interpretation biases in a simple, spoken task. Thus, we cannot explain variation in responses in terms of reading skill, in that participants did not need to read to do the task. This finding goes beyond existing evidence for exposure effects, which has mostly focused on either vocabulary or grammatical development (Anderson, Farmer, Goldstein, Schwade, & Spivey, 2011; Fernald & Marchman, 2011; Hoff, 2003; Montag & MacDonald, 2015). Instead, our findings address a very specific process in discourse understanding, namely the ability to understand ambiguous pronouns. This paper is the first to demonstrate that children who read more tend to follow the linguistic context more – and specifically the subject bias – compared to children who read less. Thus, this finding suggests that the type and/or quality of discourse exposure may play a key role in the development of strategies used to understand ambiguous pronouns.

Acknowledgements. We are grateful to the children and parents who participated in our study. We are extremely thankful to the Morehead Planetarium and Science Center, the Orange County library, and the Durham County library for allowing us to recruit through their programs. Thank you to Fly Leaf Books for donating books for the children who participated. Thank you to Brianna Torres, Kirsten Bubak, and Liz Reeder for assisting in the recruitment and administration of the task. This project was partially supported by a Stephenson-Lindquist grant from the Dept. of Psychology and Neuroscience at UNC Chapel Hill.

References

- Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, 40, 278–89.
- Almor, A. (1999). Definite noun-phrase anaphora and focus. *Psychological Review*, 106, 748–65.
- Anderson, S. E., Farmer, T. A., Goldstein, M., Schwade, J., & Spivey, M. J. (2011). Individual differences in measures of linguistic experience account for variability in the sentence processing skill of five-year-olds. In I. Arnon & E. V. Clark (Eds.), *Experience, variation, and generalization: learning a first language* (pp. 203–21). Amsterdam: John Benjamins.
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Arnold, J. E. (1998). *Reference form and discourse patterns*. Doctoral dissertation, Stanford University.

- Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, 31(2), 137–62.
- Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistics Compass*, 4(4), 187–203.
- 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, 527–65.
- Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course for pronoun resolution from eyetracking. *Cognition*, 76(1), B13–B26.
- Arnold, J. E., Nadig, A., Bennetto, L., & Diehl, J. J. (2009). Pronoun comprehension and production in children with and without autism. Poster; CUNY Conference on Human Sentence Processing, Davis, CA.
- Arnold, J. E., Strangmann, I., Hwang, H., & Zerkle, S. (2018b). Reference frequency: What do speakers tend to talk about? Technical Report #2. UNC Language Processing Lab, Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill.
- Arnold, J. E., Strangmann, I., Hwang, H., Zerkle, S., & Nappa, R. (2018a). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, 102, 41–54.
- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62, 875–90.
- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, 20, 395–418.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the Eyes' test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–51.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, 68(3), 255–78.
- Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, 8, 492–99.
- Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, 30, 359–75.
- Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States part I: variations by age, ethnicity, and poverty status. *Child Development*, 72, 1844–67.
- Bright, R. (1963). *Georgie and the Robbers*. New York: Doubleday and Company.
- Cannon, J. (1993). *Stellaluna*. San Diego, CA: Harcourt Brace Jovanovich.
- Cartmill, E. A., Armstrong, B. I., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *PNAS Proceedings of The National Academy of Sciences of the United States of America*, 110, 11278–83.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–56). New York: Academic Press.
- Crawley, R. J., & Stevenson, R. J. (1990). Reference in single sentences and in texts. *Journal of Psycholinguistic Research*, 19(3), 191–210.
- Csibra, G., & Volein, Á. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, 26, 1–11.
- Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: a quick measure of reading experience. *Journal of Educational Psychology*, 82(4), 733–40.
- Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, 83, 264–74.
- De Ayala, R. J. (2013). *The theory and practice of item response theory*. New York: Guilford Publications.
- Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, 17, 270–81.
- Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., *et al.* (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language and Hearing Research*, 42, 1432–43.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272–99.

- Farmer, T. A., Fine, A. B., Misyak, J. B., & Christiansen, M. H. (2017). Reading span task performance, linguistic experience, and the processing of unexpected syntactic events. *Quarterly Journal of Experimental Psychology*, 70, 413–33.
- Fernald, A., & Marchman, V. A. (2011). Causes and consequences of variability in early language learning. In I. Arnon, & E. V. Clark (Eds.), *Experience, variation, and generalization: learning a first language* (pp. 181–202). Amsterdam: John Benjamins.
- Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37(3), 578–91.
- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: active versus passive representations. *Journal of Memory and Language*, 56, 357–83.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998.
- Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: the advantage of first mention. *Journal of Memory and Language*, 27, 699–717.
- Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science* 12, 81–7.
- Goodrich Smith, W., & Hudson Kam, C. L. (2015). Children's use of gesture in ambiguous pronoun interpretation. *Journal of Child Language*, 42, 591–617.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311–47.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Paul H. Brookes.
- Hartshorne, J. K., Nappa, R., & Snedeker, J. (2015a). Development of the first-mention bias. *Journal of Child Language*, 42(2), 2–25.
- Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015b). The causes and consequences explicit in verbs. *Language, Cognition, and Neuroscience*, 30, 716–34.
- Heath, S. B. (1983). *Ways with words: language, life, and work in communities and classrooms*. London: Cambridge University Press.
- Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74, 1368–78.
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental Psychology*, 49(1), 4–14.
- Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, 19, 603–29.
- Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of Pediatric Psychology*, 27, 87–96.
- Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: a longitudinal analysis. *Journal of Educational Psychology*, 100(2), 252–71.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, 27, 236–48.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input at home and at school: relation to child syntax. *Cognitive Psychology*, 45, 337–74.
- James, A. N., Fraundorf, S., Lee, E., & Watson, D. (2018). Individual differences in syntactic processing: Is there evidence for reader–text interactions? *Journal of Memory and Language*, 102, 155–81.
- Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: contrasting the first-mention and subject preference accounts. *Psychological Science*, 16, 260–4.
- Kaiser, E. (2011). Focusing on pronouns: consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes*, 26(10), 1625–66.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. (2008). Coherence and coreference revisited. *Journal of Semantics Special Issue on Processing Meaning*, 25, 1–44.
- Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, 39, 1–37.

- Koster, C., Hoeks, J., & Hendriks, P.** (2011). Comprehension and production of subject pronouns: evidence for the asymmetry of grammar. In A. Grimm, A. Müller, C. Hamann, & E. Ruigendijk (Eds), *Production-comprehension asymmetries in child language* (pp. 99–122). Berlin: De Gruyter Mouton.
- Lloyd, P., Mann, S., & Peers, I.** (1998). The growth of speaker and listener skills from five to eleven years. *First Language, 18*, 81–104.
- McDonald, J. L., & MacWhinney, B.** (1995). The time course of anaphor resolution: effects of implicit verb causality and gender. *Journal of Memory and Language, 34*, 543–66.
- Meyer, M., & Baldwin, D. A.** (2013). Pointing as a socio-pragmatic cue to particular vs. generic reference. *Language Learning and Development, 9*, 245–65.
- Moll, H., & Tomasello, M.** (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science, 7*, F1–F9.
- Montag, J. L., & MacDonald, M. C.** (2015). Text exposure predicts spoken production of complex sentences in eight and twelve year old children and adults. *Journal of Experimental Psychology: General, 144*(2), 447–68.
- Moore, M., & Gordon, P. C.** (2015). Reading ability and print exposure: item response theory analysis of the author recognition test. *Behavior Research Methods, 47*(4), 1095–109.
- Morford, M., & Goldin-Meadow, S.** (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language, 19*, 559–80.
- Morisset, D., Barnard, K., Greenberg, M., Booth, D., & Spieker, S.** (1990). Environmental influences on early language development: the context of social risk. *Development and Psychopathology, 2*, 127–49.
- Namy, L. L., Campbell, A. L., & Tomasello, M.** (2004). The changing role of iconicity in non-verbal symbol learning: a U-shaped trajectory in the learning of arbitrary gestures. *Journal of Cognition and Development, 5*, 37–57.
- Nappa, R., & Arnold, J. E.** (2014). The road to understanding is paved with the speaker's intentions: cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology, 70*, 58–81.
- Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S.** (2013). Infants understand the referential nature of human gaze but not robot gaze. *Journal of Experimental Child Psychology, 116*(1), 86–95.
- Orr, A. J.** (2003). Black-White differences in achievement: the importance of wealth. *Sociology of Education, 76*, 281–304.
- Pykkönen, P., Matthews, D., & Järvikivi, J.** (2010). Three-year-olds are sensitive to semantic prominence during online spoken language comprehension: a visual world study of pronoun resolution. *Language and Cognitive Processes, 25*, 115–29.
- Rey, H. A.** (1969). *Curious George*. Boston, MA: Houghton Mifflin.
- Rohde, H., & Kehler, A.** (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience, Special Issue on Production of Referring Expressions: Models and Empirical Data, 29*, 912–27.
- Rowe, M. L.** (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*, 185–205.
- Rowe, M. L.** (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*, 1762–74.
- Russell, A. E., Ford, T., Williams, R., & Russell, G.** (2016). The association between socioeconomic disadvantage and attention deficit/hyperactivity disorder (ADHD): a systematic review. *Child Psychiatry Human Development, 47*, 440–58.
- Sobel, D. J.** (1963). *Encyclopedia Brown, Boy Detective: The Case of the Scattered Cards*. New York: Puffin Books.
- Song, H., & Fisher, C.** (2005). Who's 'she'? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language, 52*(1), 29–57.
- Song, H., & Fisher, C.** (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua, 117*, 1959–87.
- Stanfield, C., Williamson, R., & Özçalışkan, Ş.** (2014). How early do children understand gesture-speech combinations with iconic gestures? *Journal of Child Language, 41*(2), 462–71.
- Stanovich, K. E., & West, R. F.** (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*, 402–33.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W.** (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505.

van Rij, J., van Rijn, H., & Hendriks, P. (2011). WM load influences the interpretation of referring expressions. In F. Keller & D. Reitter (Eds.), *Proceedings of the 2nd workshop on Cognitive Modeling and Computational Linguistics* (pp. 67–75). Portland, OR: Association for Computational Linguistics.

van Rij, J., van Rijn, H., & Hendriks, P. (2013). How WM load influences linguistic processing in adults: a computational model of pronoun interpretation in discourse. *Topics in Cognitive Science*, 5(3), 564–80.

Vernon-Feagans, L., Hammer, C. S., Miccio, A., & Manlove, E. (2001). Early language and literacy skills in low-income African American and Hispanic children. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 192–210). New York: Guilford Press.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250–71.

Woodward, A. L. (2003). Infants’ developing understanding of the link between looker and object. *Developmental Science*, 6, 297–311.

Yow, W. Q. (2013). Monolingual and bilingual children’s use of gestures and grammatical agreement in pronoun interpretation. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *BUCLD 37 Proceedings* (pp. 482–94). Somerville, MA: Cascadia Press.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLOS One*, 8, e79659. doi:10.1371/journal.pone.0079659

Appendix A

The following parent questionnaire was based on Hood, Conlon, and Andrews (2008).

Parent Questionnaire: Reading at Home

For the next three questions, please think about three time periods:

- a) at the time when your child was 3 years old
- b) at the time when your child was 5 years old
- c) over the last 3 months

For each period of time, please make your best estimate.

1. In a typical week, how many hours (on average) do you or someone else read to your child?

LAST 3 MO.	AGE 5	AGE 3	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	rarely or never
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1 hour per week (about 15 minutes, 4 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2 hours (about 30 minutes a day, 4 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3 hours (about 30 minutes a day, 6 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4–6 hours (about 40–60 minutes a day, 6 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7 hours (about an hour a day, every day)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8 hours + (more than an hour a day)

2. Not including school assignments, in a typical week how often does your child read to him/herself or look at books, magazines, or other printed material?

LAST 3 MO.	AGE 5	AGE 3	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	rarely or never
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1 hour per week (about 15 minutes, 4 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2 hours (about 30 minutes a day, 4 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3 hours (about 30 minutes a day, 6 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4–6 hours (about 40–60 minutes a day, 6 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7 hours (about an hour a day, every day)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8 hours + (more than an hour a day)

3. Not including school assignments, in a typical week how often does your child listen to books on CD, tape, or other media?

LAST 3 MO.	AGE 5	AGE 3	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	rarely or never
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1 hour per week (about 15 minutes, 4 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2 hours (about 30 minutes a day, 4 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3 hours (about 30 minutes a day, 6 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4–6 hours (about 40–60 minutes a day, 6 days a week)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7 hours (about an hour a day, every day)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8 hours + (more than an hour a day)

4. How much does your child enjoy hearing someone else read to him/her?
 loves it likes it neither likes nor dislikes it not very enthusiastic about it

5. How much does your child enjoy reading to him/herself?
 loves it likes it neither likes nor dislikes it not very enthusiastic about it

6. Is there anything else you would like to tell us about your child's reading habits?

Appendix B

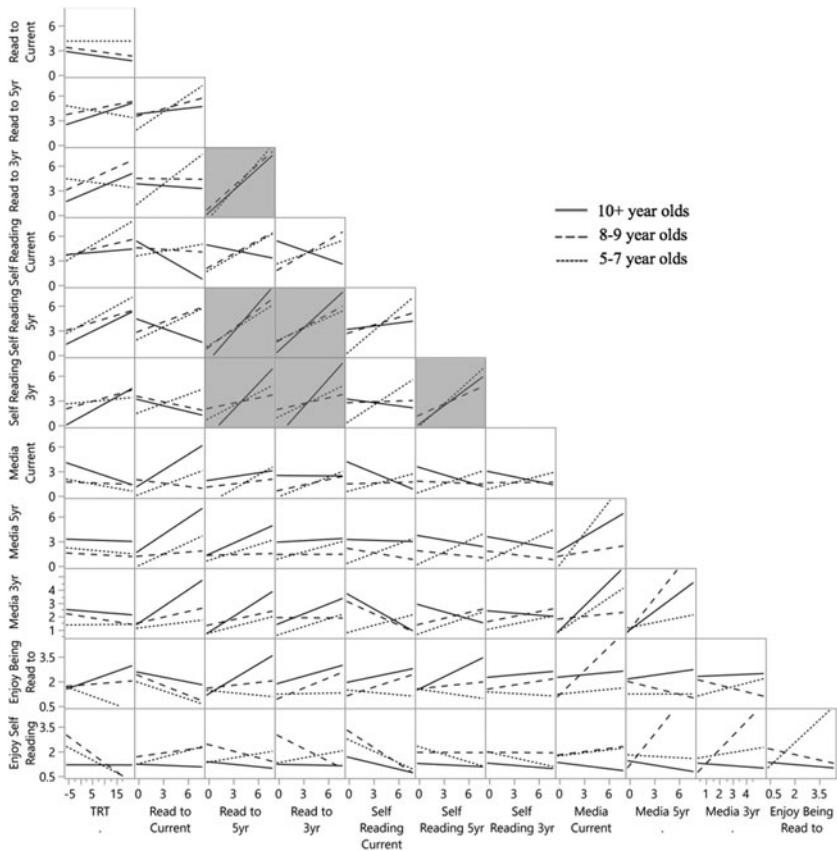


Figure S1. Matrix with regression lines for all items in parental questionnaire. Lines of best fit reflect the positive, negative, or null association across items. Statistical significance is not illustrated in this figure. The six shaded regions are the ones used in our exploratory factor analysis.

Cite this article: Arnold JE, Castro-Schilo L, Zerkle S, Rao L (2019). Print exposure predicts pronoun comprehension strategies in children. *Journal of Child Language* 46, 863–893. <https://doi.org/10.1017/S0305000919000102>