Development of the first-mention bias

(Article begins on next page)

Journal of Child Language

Development of the first-mention bias

JOSHUA K. HARTSHORNE, REBECCA NAPPA and JESSE SNEDEKER

How to cite this article:
JOSHUA K. HARTSHORNE, REBECCA NAPPA and JESSE SNEDEKER
Development of the first-mention bias . Journal of Child Language, Available on
CJO 2014 doi:10.1017/S0305000914000075

Request Permissions : Click here

Development of the first-mention bias*

JOSHUA K. HARTSHORNE, REBECCA NAPPA AND
JESSE SNEDEKER

Harvard University

ABSTRACT

In many contexts, pronouns are interpreted as referring to the character mentioned first in the previous sentence, an effect called the 'first-mention bias'. While adults can rapidly use the first-mention bias to guide pronoun interpretation, it is unclear when this bias emerges during development. Curiously, experiments with children between two and three years old show successful use of order of mention, while experiments with older children (four to five years old) do not. While this could suggest U-shaped development, it could also reflect differences in the methodologies employed. We show that children can indeed use first-mention information, but do so too slowly to have been detected in previous work reporting null results. Comparison across the present and previously published studies suggests that the speed at which children deploy first-mention information increases greatly during the preschool years.

INTRODUCTION

Pronouns have no fixed reference; rather, reference depends on context:

(1) a. Jane Austen is my favorite author. **She** wrote many popular books.
   b. Ursula LeGuin is my favorite author. **She** wrote many popular books.

This is in contrast to proper names, which do not depend on context:

(2)   a. Jane Austen is an author. **Jane Austen** wrote many books.
      b. Ursula LeGuin is an author. **Jane Austen** wrote many books.

Third person pronouns frequently co-refer with the subject of the previous sentence. For example, Arnold (1998) found that third person subject pronouns co-referred with the previous sentence's subject in 64% of cases in a corpus of children's books. Adult comprehenders are sensitive to this pattern and typically expect pronouns to co-refer with the previous subject, even in the absence of additional clues to reference or when alternate interpretations are plausible (Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000; Corbett & Chang, 1983; Crawley & Stevenson, 1990; Crawley, Stevenson & Kleinman, 1990; Gordon, Grosz & Gilliom, 1993; Gordon & Scearce, 1995; Järvikivi, van Gompel, Hyona & Bertram, 2005; Kaiser & Trueswell, 2008; Smyth, 1994; Yang, Gordon, Hendrick & Hue, 2003). Thus, in (3), most adults prefer that *she* refer to Jane Austen, not Agatha Christie.

(3)   Jane Austen was born long before Agatha Christie. She wrote many books.

In English, the subject of a sentence is also almost always the first-mentioned noun. Consequently, this bias has typically been called the 'first-mention bias', a term we adopt here. Note that additional research, particularly work in languages where order-of-mention and subject-hood are more easily de-confounded, has suggested that subject-hood and order-of-mention each play distinguishable roles (Gordon & Chan, 1995; Järvikivi *et al*., 2005; Kaiser & Trueswell, 2008; Crawley & Stevenson, 1990).

Many researchers have conceptualized the first-mention bias as a global bias that applies in all contexts. However, pronoun interpretation biases are at least partly a function of discourse structure and consequently vary depending the content of the sentences and their connection (see especially Kehler, 2002; Kehler, Kertz, Rohde & Elman, 2008). While many discourse structures support first-mention biases, others do not:

(4)   Jane Austen liked Agatha Christie because she . . .

Here, *she* likely refers to Agatha Christie, the second-mentioned character.

For the present, this issue is orthogonal to our main point. Our study and literature review focuses on children's processing of sentences which, in adults, reliably lead to first-mention biases. We turn to other contexts in the 'Discussion'.

The development of the first-mention bias

While several studies on the development of the first-mention bias have been reported, results are mixed. While results of a number of experiments suggest that even very young children are sensitive to the first-mention bias (Pyykkönen, Matthews & Järvikivi, 2010; Song & Fisher, 2005, 2007), results of two others indicate that they are not (Arnold, Brown-Schmidt & Trueswell, 2007). Below we consider three plausible explanations for the divergence in these findings.

Statistical error. Perhaps the simplest explanation is that either the findings that children are sensitive to the first-mention bias or the findings that they are not are in error. Although more experiments have shown positive results (five) than negative results (two), simple vote-tallying may not work, since this evidence must be interpreted in the context of how many false positives and false negatives one expects to find in the literature, which is an underdetermined and controversial question.

There is an overwhelming bias in psychology against publishing null results (for review, see Hartshorne & Schachner, 2012). Thus, it is much harder to publish false negatives than false positives, increasing the false positive to false negative ratio in the literature. In contrast, meta-analyses indicate that the typical psychology experiment is under-powered, with less than a 50% chance of rejecting the null hypothesis even when the null hypothesis is in fact false (Bakker, van Dijk & Wicherts, 2012; Hartshorne & Schachner, 2012). Thus, the null hypothesis is far more likely to be falsely accepted (>50%) than falsely rejected (<5%). These and other factors make it difficulty to determine the likely ratio of false positives to false negatives in psychology.

There is, however, good reason to suspect that one of the results indicating insensitivity to the first-mention bias in children is a false negative. In control trials for Experiment 1 of Arnold *et al*. (2007), pronoun reference was disambiguated by gender (*Puppy* [male] *is having lunch with Froggy* [female]. *She wants some milk*.), and thus success in interpreting the pronoun should have been at ceiling. Nonetheless, children and adults each chose the correct animal only 82% of the time (compared with chance performance of 50%), suggesting deep confusion about the task, leading them to produce either random or systematically incorrect results some portion of the time, noise which could wash out the effect of interest.

U-shaped development. The experiments showing sensitivity to the first-mention bias tested younger children (2;6–3;5) than the experiments showing insensitivity (3;5–5;9), raising the intriguing possibility of U-shaped development. While language skills generally improve throughout development, there are aspects of development that show systematic dips in performance. For instance, success at interpreting passive sentences drops

during the preschool years before eventually recovering and reaching adult performance (Bever, 1970; Maratsos, 1974). Likewise, children initially make few if any over-regularization errors for past tense forms (*goed, holded, sleeped*); if they produce a past tense form at all, they produce the correct irregular form (*went, held, slept*). Over-regularization errors increase at ages three to four, and then slowly decline thereafter (Cazden, 1968; Ervin, 1964; Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992). U-shaped developmental trajectories are typically taken as indicating that children initially learn to process, e.g., passives or irregular past tense verbs one-by-one in an essentially item-based fashion. Then, at some point, children discover a pattern elsewhere in the language that they overgeneralize to exceptional cases. In the case of passives, they discover the dominant active sentence pattern in which the first verbal argument is usually the agent. In the case of the past tense, they discover that most verbs form their past tense by adding -*ed*, and they then overapply this to irregulars. This overgeneralization leads to a spurt of errors, which are only overcome as the system is more fully acquired.

U-shaped development for the first-mention bias could have a similar explanation: by two and half years of age, children may discover that pronouns typically co-refer with the first-mentioned entity from the previous sentence, leading to high performance on sentences in which adults also show first-mention biases. However, as they grow older they begin to notice that there are discourse contexts in which there is a systematic bias for subject pronouns to refer to the object of the previous sentence (see (4), above). To capture these less frequent patterns, children would have to abandon their global first-mention bias in favor of a set of narrower – but more precise – generalizations. Initially, however, they may be unsure about which types of sentence are governed by which generalizations. This could lead older preschoolers, like those in the Arnold *et al*., studies (2007), to perform at chance in contexts where both younger children and adults show a systematic first-mention bias.

Processing speed. Strictly speaking, Arnold *et al*.'s (2007) Experiment 2 did not show that children fail to use first-mention information to guide pronoun interpretation, only that they fail to use that information within 650 ms after pronoun onset. Like all the experiments discussed above, this experiment employed the visual-world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995), a form of preferential looking (Fagan, 1971; Fantz, 1958; Spelke, 1985). Children listened to stories like (5) while looking at an illustration of the story.

(5)   Donald is bringing some mail to Mickey, while a big rainstorm is beginning. He's carrying an umbrella, and it looks like they're both going to need it.

TABLE 1. *Visual-world studies of the first-mention bias in children and adults*

Study	Age	Length of ambig. region	1st window w/ sig. 1st-men. bias
Song & Fisher, 2007	2;6	3700 ms	3000–4000 ms
Song & Fisher, 2005 (Exp. 2)	3;0	3700 ms	3000–4000 ms
Song & Fisher, 2005 (Exp. 3)	3;0	3700 ms	3000–4000 ms
Song & Fisher, 2005 (Exp. 4)	3;0	3700 ms	1000–2000 ms
Pyykkönen et al., 2010	3;5	>4000 ms	1240–1760 ms
Arnold et al., 2007 (Exp. 1)	4;0–5;9	N/A[1]	none
Arnold et al., 2007 (Exp. 2)	4;0–5;9	650 ms	none
Arnold et al., 2000	adult	650 ms	400–600 ms
Järvikivi et al., 2005	adult	>1000 ms	480–900 ms[2]
Song & Fisher, 2005	adult	3700 ms	1000–2000 ms[3]

NOTES:
[1] Only the first 3000 ms after pronoun onset were analyzed.
[2] Järvikivi et al. (2005) crossed grammatical role and order of mention in Finnish. The effect of subjecthood, independent of order of mention, appeared at 480–690 ms. The effect of order of mention, independent of subjecthood, appeared 690–900 ms after pronoun onset.
[3] Note that the effect trended towards significance ($p = \cdot 09$) in the 0–1000 ms time window.

The dependent measure was whether participants looked more at Donald (the first-mentioned character) or at Mickey in the period after the onset of the pronoun (*he's*) and before the disambiguating word (*umbrella*; in the illustration, Donald carries an umbrella and Mickey does not). Critically, subsequent looks to Donald cannot be attributed to a first-mention bias because they may be due to the disambiguating information.

While all the looking-time studies employed a similar method, restricting analyses to a brief ambiguous time window, the studies in which children showed sensitivity to first-mention all employed significantly longer ambiguous regions, giving children more time (Table 1). Moreover, in all of these studies, the first evidence of sensitivity to first-mention information emerged only well after 650 ms post pronoun onset.

Is it possible that children had insufficient time to show their sensitivity to first-mention in Arnold *et al.*'s (2007) Experiment 2? Arnold *et al.*, considered and rejected this hypothesis for two reasons. First, in an earlier study using the same design (Arnold *et al.*, 2000), adults had successfully demonstrated sensitivity to first-mention within the time allotted. This could not be explained by generally slower processing on the part of the children because they were just as fast in a separate condition in which the pronoun was disambiguated by gender. Second, in Arnold and colleagues' additional experiment (Experiment 1), the pronoun was globally ambiguous (*Puppy* [male] *is having lunch with Panda Bear* [male]. *He wants some milk*.). Although children had ample time to respond, at no point were they more likely to look at the first-mentioned character. When asked at the end of the

5

trial to explicitly resolve the pronoun (e.g. answering the question "Who wants milk?"), they chose at chance rates.

There are good reasons to revisit both these lines of argumentation. First, in all five experiments in which children did show sensitivity to first-mention, the effect emerged well after 650 ms post pronoun onset (Table 1). Had children in those studies been restricted to 650 ms, those studies would have concluded – like Arnold *et al*. (2007) – that children are insensitive to first-mention. In fact, 650 ms may be a tight time limit for adults: in both of the other experiments involving adults, the first significant time window extended beyond 650 ms (Table 1). Moreover, it may well be that although five-year-olds are as fast as adults at disambiguating pronouns based on gender, they are nonetheless slower at using first-mention. Differential effects of development on processing speed across different linguistic stimuli have been observed elsewhere. For instance, while the speed of processing nouns increases during the second year of life and then remains stable during the third, the speed of processing adjectives continues to develop (Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998; Fernald, Perfors & Marchman, 2006; Fernald, Thorpe & Marchman, 2010).

Second, as noted above, confidence in the results of Arnold *et al*.'s (2007) Experiment 1 – the experiment involving globally ambiguous pronouns – is weakened by the fact that in another condition, both adults and children preferred to resolve the pronoun to a character of the wrong gender 1/5 of the time. Since we would not want to conclude, at least in the case of adults, that participants think that *she* refers to a male 20% of the time, it is not clear that we should interpret their results in the ambiguous-pronoun condition as directly reflecting sensitivity (or lack thereof) of a first-mention bias.

THE EXPERIMENT

In order to contrast the STATISTICAL ERROR, U-SHAPED DEVELOPMENT and PROCESSING SPEED hypotheses, we conducted a Visual World Paradigm study employing globally ambiguous sentences with five-year-olds and adults. An example image is shown in Figure 1. Examples of stories from each of four conditions are shown in Table 2.

This design provides an off-line measure of pronoun resolution (which character is pointed to), mitigating worries about artificial time limits on children's responses. Note that with the exception of the unambiguous gender condition, these utterances did not contain disambiguating words, and thus the time over which a first-mention bias could emerge is unlimited. If the U-SHAPED DEVELOPMENT HYPOTHESIS is correct – or if previous studies showing successful use of first-mention by children are false positives –five-year-olds should fail to show a first-mention bias in the FIRST-MENTION
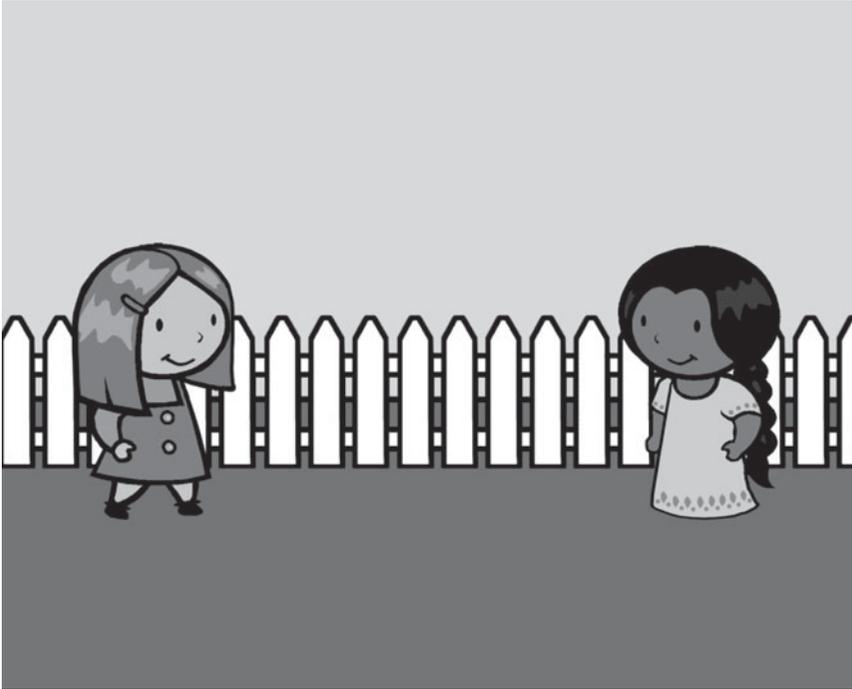
Fig. 1. An example image from the experiment, which participants viewed while hearing passages like "Emily went to school with Hannah. She read ten books." Emily is on the left and Hannah is on the right. Children were taught characters' names prior to the experiment.

CONDITION in either their eye-movements or their points. If the PROCESSING SPEED HYPOTHESIS is correct, the first-mention bias will appear in both measures, but looks to the first-mentioned character should not rise above chance until well after the time when they do in adults – and after the 650 ms ambiguous window present in the second Arnold *et al*. (2007) experiment.

As in the Arnold studies, the GENDER CONDITION serves as a baseline, revealing how quickly and accurately participants can disambiguate pronouns based on gender alone. The character with matching gender was equally likely to be the first-mentioned or last-mentioned character.

We included two additional conditions that may help us better understand what changes in pronoun processing over development. One possible explanation for children's slow and/or poor performance on first-mention tasks is that they have difficulty re-accessing their representation of the first-mentioned character after a delay. To test this hypothesis, we included the REPEATED-MENTION (SHORT) and REPEATED-MENTION (LONG) conditions,

TABLE 2. *Example stimuli from the present experiment. The critical pronoun is underlined*

Condition	Example story
First-mention	Emily ate dinner with Hannah. <u>She</u> skipped her salad and only ate dessert. Can you point to her?
Repeated-mention (short)	Emily and Hannah are going to Disneyland. Emily has never been to Disneyland. <u>She</u> is really excited about going to Disneyland. Can you point to her?
Repeated-mention (long)	Emily and Hannah are going to Disneyland. Emily has never been to Disneyland. Disneyland has lots of fun activities. It also has great food. <u>She</u> is really excited about going to Disneyland. Can you point to her?
Gender (unambiguous)	Emily played baseball with Michael. <u>S/he</u> hit five homeruns. Can you point to her?

in which one character is made particularly salient by being mentioned first and mentioned repeatedly. The length of time between the pronoun and the most recent mention of the target character was manipulated by inserting two additional filler sentences, where neither character is mentioned, into the LONG condition (Table 2). Alternately, children may have no difficulty re-accessing the most recently mentioned character in the absence of a competing referent, in which case performance in these conditions should be strong and fast.

METHOD

Participants

Forty five-year-olds (4;11–6;1; $M = 5;7$) and thirty-six adults participated. Nine additional adults and three additional children were excluded for excessive track loss (track loss in remaining participants: 2%–49%, $M = 20$%). More aggressive thresholds resulted in qualitatively similar results. All participants were native English speakers. Parents did not disclose any developmental deficits or disabilities and no evidence of such disabilities was observed during testing.

Materials

Four novel cartoon characters (two male, two female) were created. Visual scenes like Figure 1 were created, with two of the characters standing equidistant on opposite sides of a midline. Characters were not engaged in any activities, and the background was simple and did not disambiguate the referent of the pronoun (a fence, a tree, a rock, a house, etc.).

Each participant received eight GENDER CONDITION trials and four trials in each of the other conditions. In all conditions, which of the four characters

was first-mentioned was counterbalanced within and between participants. Each of the eight REPEATED-MENTION CONDITION stories came in a SHORT and LONG variant; which variant was used was counterbalanced across subjects. Similarly, in the GENDER CONDITION we counterbalanced which character (first-mentioned or second-mentioned) matched the pronoun. Due to a recording error, half of the participants were given the wrong point- ing instruction ("Can you point to her") for one of the eight unambiguous sentences (the gender of the pronoun was incorrect). As the error occurred over four seconds after the onset of the critical pronoun, it did not affect the eye-tracking analyses; however, the pointing response for that item was excluded from the off-line response analyses for those participants.

The counterbalancing discussed above produced four lists. Stimuli were placed in a pseudo-random order such that no condition appeared more than twice in a row. The same order for the stories was used on all four lists, though with the variants of each story counterbalanced as described above. Four additional lists were created by reversing the order of the stimuli, for a total of eight lists. Four to six children and three to six adults were tested on each list.

Sentences were recorded by a female native English speaker using child-directed prosody but with no differential stress on any of the noun phrases, including the pronouns. The average time elapsed between onset of the pronoun and onset of the most recently mentioned character was 1156 ms in the first-mention condition (min: 1071 ms), 7566 ms in the repeated-mention (short) condition (min: 6202 ms), 13709 ms in the repeated-mention (long) condition (min: 11813 ms), and 1340 ms in the unambiguous gender condition (min: 1137 ms).

Procedure

Participants were familiarized with the characters and drilled on their names until they could name each character rapidly and without hesitation. They were then told they would look at pictures and listen to accompanying stories. In each story, they would be asked to point to someone. If they were not sure who that person was, they should make their best guess. Stimuli were presented on a Tobii T60 desktop eye-tracker, which also recorded eye-movements. The picture appeared on the screen one second before the sound file started playing. Off-line responses (pointing) were recorded by the experimenter. The entire procedure, including calibration of the eye-tracker, took approximately 15 minutes.

RESULTS

Both off-line (pointing) and on-line results were analyzed using binomial mixed-effects models in R (Bates & Sarkar, 2007;

Percentage Choosing Target
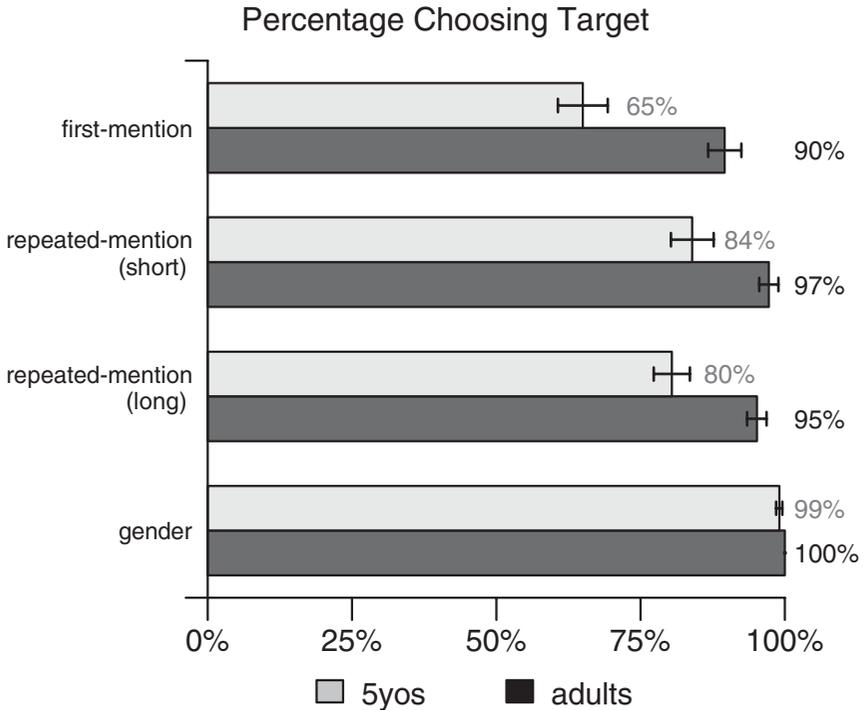
Fig. 2. Off-line (pointing) results with standard errors. Note that while the difference between adults and children appears to be greatest in the first-mention condition, this is an artifact of binomial distributions: differences near 50% are less meaningful than differences near 0% or 100% (for review, see Jaeger, 2008).

R-development-core-team, 2005), with subjects and sentence templates as random effects. In no case did including random slopes for initial fixation improve model fit, nor was the pattern of significance different for models with or without random slopes. Thus, only the simpler random-intercepts-only models are reported.

Off-line (pointing) results

Participants' off-line responses are shown in Figure 2. The weakest performance was by children in the first-mention condition, so that was used as the reference cell for the mixed-effects regression. The intercept was positive and significant ($z = 3 \cdot 4$, $p < \cdot 001$), indicating that, even in this condition, participants were more likely to point to the target character than would be predicted by chance. Performance was significantly higher in the repeated-mention short ($z = 4 \cdot 1$, $p < \cdot 0001$), repeated-mention long

Fig. 3. Results for first-mention condition in terms of percent looks to target (left panel) and switch preference (percent switched from distractor to target minus percent switched from target to distractor; right panel).

($z = 3 \cdot 1$, $p = \cdot 002$), and unambiguous conditions ($z = 6 \cdot 0$, $p < \cdot 0001$). Across conditions, adults were more likely to point to the target character than were children ($z = 4 \cdot 5$, $p < \cdot 00001$). There were no significant interactions between subject group and condition (zs < 1).
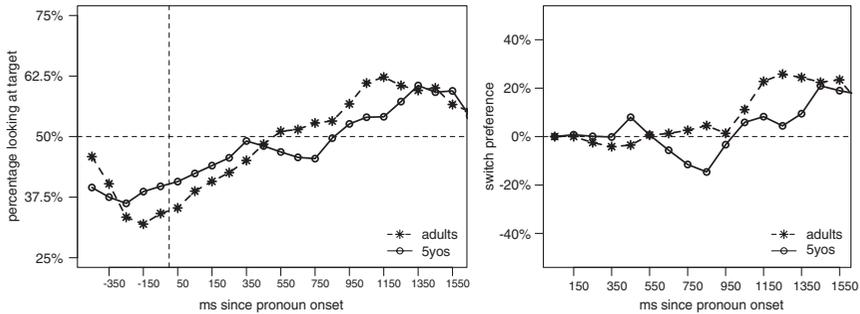
On-line (eye-tracking) results

Throughout, direction of eye gaze was analyzed by a midline split: participants looking to the left of midline were considered to be fixating the character on the left. Note that the characters were equidistant from the midline.

First-mention condition. The off-line (pointing) results indicate that five-year-olds are indeed sensitive to first-mention information. According to the processing speed hypothesis, children failed to show such sensitivity in Arnold *et al*.'s (2007) Experiment 2 because they did not have enough time to demonstrate that sensitivity (possibly because they are systematically slower than adults). To test this hypothesis, we turned to the eye-tracking results.

Fixations were binned into successive 100 ms windows relative to pronoun onset. If a saccade occurred during a window, the participant was credited as looking at whichever character they were looking at for the majority of that window. Results are shown in Figure 3 (left panel). Children were not more likely to look at the target character until 1400–1500 ms post pronoun onset. This is well after the 650 ms cut-off in Arnold *et al*. (2007) but well within the time allowed in other studies. Adults were slightly faster, showing a significant first-mention effect by 1100–1200 ms post pronoun onset.

Note, however, that at pronoun onset, both children and adults were significantly more likely to be looking at the second-mentioned (non-target) character than the first-mentioned (target) character (Looks to

first-mentioned: Adults = 35%, $p = \cdot 001$; five-year-olds = 41%, $p = \cdot 02$), likely because the second-mentioned character had been mentioned 1071–2025 ms earlier (median = 1156 ms).

In principle, this baseline effect could either result in artificially slow responses (looks to target must overcome the baseline effect) or artificially fast responses (because people tend to look away from whatever they have been looking at, these novelty-driven looks away are indistinguishable from pronoun-driven looks to the target). If these baseline effects lead to artificially fast responses, that would only support the processing speed hypotheses that much more. Thus, we considered the possibility that they led to artificially slow results.

To correct for the baseline effect, we analyzed shifts off the character being fixated at pronoun onset (Fernald *et al*., 1998). We divided the trial in terms of whether the participant was initially fixating the target or distractor during the first 100 ms following pronoun onset. During each subsequent 100 ms bin, we coded whether the participant was still fixating the originally fixated character or was now fixating the other character. If a saccade took place during a 100 ms bin, the bin was coded according to where the participant was looking during the majority of the bin. The critical question was whether the proportion of shifts from distractor to target exceeded the proportion of shifts from target to distractor, and, if so, when this occurred. In other words, when are there more switches in the 'correct' direction than in the 'wrong' direction? To answer this question, we coded each trial according to whether the participant was looking at the target or distractor at the onset of the critical word. In each subsequent time bin, we calculated the percentage of those participants who were initially fixating the distractor but were now fixating the target – that is, those who switched in the 'correct' direction – and subtracted the percentage of those participants who were initially fixating the target but were now fixating the distractor – that is, those who switched in the 'wrong' direction. Thus, a positive number indicates that more have switched in the correct direction, whereas a negative number indicates that more have switched in the wrong direction. Interestingly, this analysis led to identical results as the original analyses, with the first significant window for adults being at 1100–1200 ms and for children at 1400–1500 ms (Figure 3, right panel; Table 3).

Thus, under either analysis, children's sensitivity to the first-mention bias – demonstrated in the off-line (pointing) results – emerged too slowly in the eye-tracking results to have appeared in Arnold *et al*.'s (2007) Experiment 2, and more slowly than adults. To better understand differences in speed of processing, we turn to the other three conditions.

Repeated mention. Results for both repeated-mention conditions are shown in Figure 4. At pronoun onset, participants were already fixating the target

TABLE 3. *Adult results: significance tests (Wald's z), comparing the proportion of adults initially fixating distractor who are now fixating target against the proportion of adults initially fixating target who are now fixating distractor, analyzed separately for each condition and for each 100 ms window (e.g. '150 ms' = '100–200 ms after pronoun onset'). Significant time windows are in bold.*

|	150 ms	250 ms	350 ms	450 ms	550 ms	650 ms	750 ms	850 ms	950 ms	1050 ms	1150 ms	1250 ms	1350 ms	1450 ms	1550 ms
Gender															
z	1·30	1·43	0·91	0·38	1·53	1·90	**2·45**	**3·70**	**4·06**	**4·37**	**5·23**	**5·64**	**4·55**	**4·55**	**4·04**
p	·19	·15	·36	·70	·13	·06	**·01**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**
First-mention															
z	0·24	0·36	0·47	0·32	0·11	0·24	0·12	0·32	0·11	1·19	**2·44**	**2·60**	**2·42**	**2·26**	**2·34**
p	·81	·72	·64	·75	·91	·81	·91	·75	·91	·23	**·02**	**·009**	**·02**	**·02**	**·02**
Repeated-mention (short)															
z	0·12	1·70	**2·54**	**2·59**	**3·12**	**3·16**	**2·87**	**2·83**	**3·45**	**4·01**	**4·32**	**3·60**	**3·61**	**3·41**	**4·07**
p	·91	·09	**·01**	**·01**	**·002**	**·002**	**·004**	**·005**	**·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**
Repeated-mention (long)															
z	0·00	0·32	1·66	1·92	1·82	1·77	1·71	1·42	1·82	1·55	**1·89**	**2·74**	**3·54**	**2·64**	**2·75**
p	1·00	·75	·10	·06	·07	·08	·09	·16	·12	·06	**·006**	**<·001**	**<·001**	**<·001**	**<·001**
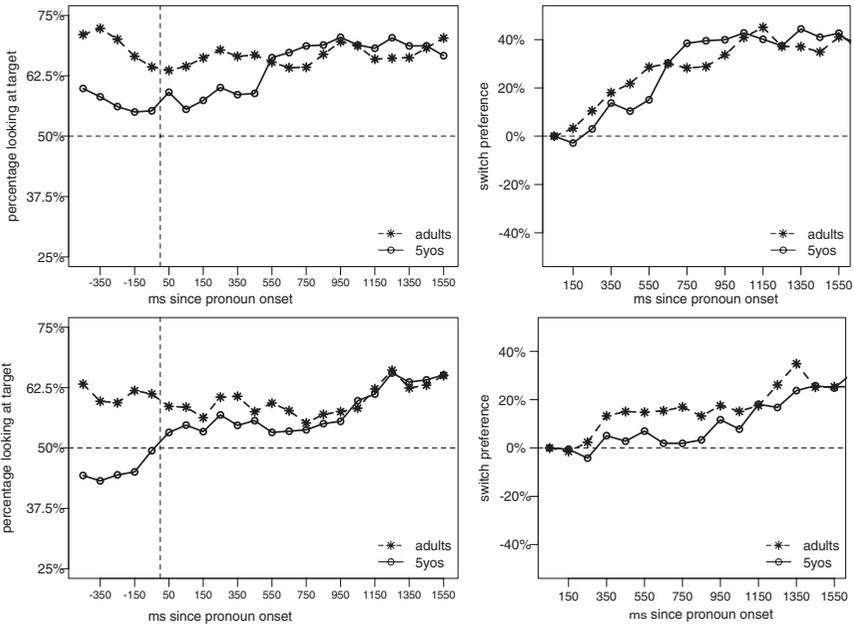
Fig. 4. Results for repeated mention short (top panels) and repeated mention long (bottom panels) in terms of percent looks to target (left panels) and switch preference (percent switched from distractor to target minus percent switched from target to distractor; right panels).

character at above-chance rates, an effect with was significant in the REPEATED-MENTION SHORT condition (adults = 64%, $p = \cdot 03$; five-year-olds = 59%, $p = \cdot 04$) though not in the REPEATED-MENTION LONG condition (adults = 59%, $p = \cdot 07$; five-year-olds = 53%, $p = \cdot 4$). This baseline effect makes it difficult to use percent-looking-to-target analyses to determine how quickly participants were able to resolve the pronoun. The switch analyses described in the previous section revealed that adults had begun to resolve the pronoun by 300–400 ms in the repeated-mention short condition. In the same time window, the five-year-olds' data is trending in the same direction ($p = \cdot 07$), though the result is not fully significant until 600–700 ms post pronoun onset. Both adults and children require 1100–1200 ms to resolve the pronoun in the repeated-mention long condition, though adults begin trending in the correct direction by 400–500 ms post pronoun onset.

Unambiguous. The GENDER trials were counterbalanced such that the target character was equally likely to be first-mentioned or second-mentioned; as such, there was no baseline bias (adults = 50%, five-year-olds = 50%). Results are depicted in Figure 5. By 600–700 ms post pronoun onset, adults showed a trend ($z = 1 \cdot 7$, $p = \cdot 095$) towards fixating the target character,
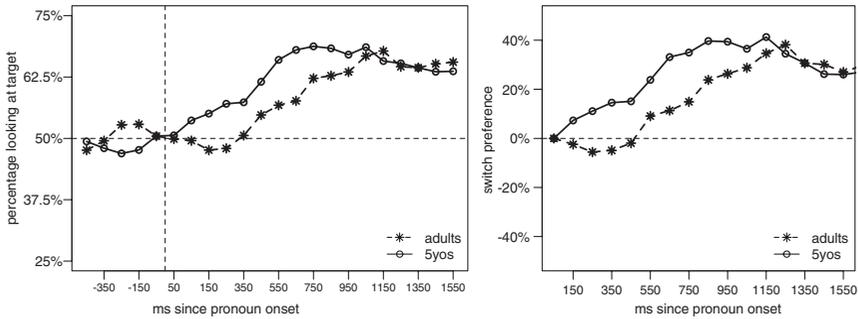
Fig. 5. Results for unambiguous condition in terms of percent looks to target (left panel) and switch preference (percent switched from distractor to target minus percent switched from target to distractor; right panel).

an effect which reached full significance by 700–800 ms ($z = 2 \cdot 4$, $p = \cdot 02$). Split analyses (as described above) showed identical results (see Table 3). For children, the effect reached significance somewhat earlier than adults (300–400 ms; $z = 2 \cdot 06$, $p = \cdot 04$), an effect which appeared even earlier in the split analyses (see Table 4).

Note that the very early effect of gender in children (100–200 ms) is somewhat surprising. We further investigated these results. In half the trials, the pronoun referred to the first-mentioned character; in the other half, the second-mentioned character. Closer inspection reveals that children were much slower to resolve the pronoun in the former case than in the latter (Figure 6, Table 5). In contrast, adults were equally fast in both cases (Figure 6, Table 6). We discuss these results below (see 'Most-recent mention bias').

DISCUSSION

We find that, like adults, five-year-olds preferred to resolve ambiguous pronouns to the first-mention character in the previous sentence. This is not due to learning during the task: five-year-olds were in fact more likely to point to the first-mentioned character in the first-mention condition during the first half of the task ($M = 70\%$) than in the task overall ($M = 65\%$). Analysis of eye-movements provides a likely explanation for the fact that Arnold and colleagues' (2007) Experiment 2 showed no effect of first-mention: the ambiguous window (650 ms) was too short. In no other study – including the present one – have children shown sensitivity to first-mention in less than 1000 ms. Thus neither statistical error nor U-shaped development are needed to explain the discrepancy between this study and those of Song and Fisher or Pyykkönen and colleagues, since there is no discrepancy.

15

TABLE 4. *Five-year-old results: significance tests (Wald's z), comparing proportion of five-year-olds initially fixating distractor who are now fixating target against the proportion of five-year-olds initially fixating target who are now fixating distractor, analyzed separately for each condition and for each 100 ms window (e.g. '150ms'='100–200 ms after pronoun onset'). Significant time windows are in bold.*

|	150 ms	250 ms	350 ms	450 ms	550 ms	650 ms	750 ms	850 ms	950 ms	1050 ms	1150 ms	1250 ms	1350 ms	1450 ms	1550 ms
Gender															
z	**2·09**	**2·37**	**2·62**	**2·52**	**3·79**	**5·20**	**5·45**	**6·09**	**6·03**	**5·67**	**6·53**	**5·51**	**4·82**	**4·12**	**4·22**
p	**·04**	**·02**	**·01**	**·01**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**
First-mention															
z	0·15	0·02	0·02	0·93	0·08	0·64	1·32	1·65	0·38	0·66	0·95	0·48	1·08	**2·38**	**2·27**
p	·88	·99	·98	·35	·94	·52	·19	·10	·71	·51	·34	·63	·28	**·02**	**·02**
Repeated-mention (short)															
z	0·68	0·49	1·82	1·21	1·74	**3·38**	**4·13**	**4·20**	**4·18**	**4·47**	**4·08**	**3·92**	**4·55**	**4·20**	**4·30**
p	·49	·63	·07	·22	·08	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**
Repeated-mention (long)															
z	0·12	0·70	0·76	0·39	0·90	0·25	0·24	0·40	1·30	0·87	**2·01**	1·81	**2·52**	**2·73**	**2·67**
p	·91	·48	·45	·69	·37	·81	·81	·69	·19	·38	**·05**	·07	**·01**	**·006**	**·008**
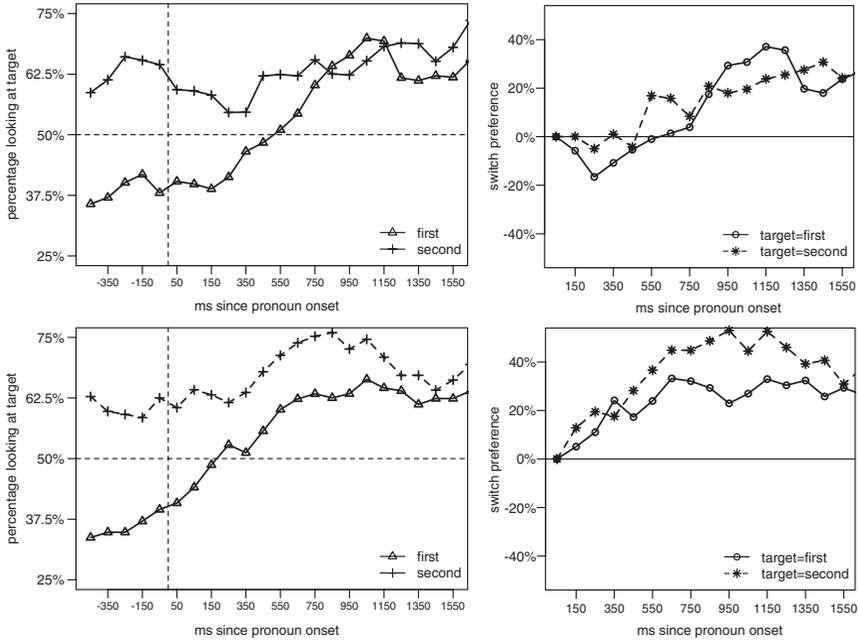
Fig. 6. Results for the unambiguous condition, with trials on which the pronoun was resolved to the first- and second-mentioned character analyzed separately. Results for adults are shown in the top panels and results for children in the bottom panels, graphed in terms of percent looks to target (left panels) and switch preference (percent switched from distractor to target minus percent switched from target to distractor; right panels).

The one remaining complication is Arnold and colleagues' global ambiguity study (2007, Experiment 1), in which there was no time limit on response and yet children nonetheless showed no evidence of a first-mention bias. However, as discussed in the 'Introduction', both the children and adults in that study also had difficulty using gender to assign pronoun reference. Since we can be reasonably confident that adults are quite adept at resolving male pronouns to males and female pronouns to females – and in fact even children were at ceiling in our own experiment – this suggests that a straightforward interpretation of the Arnold global ambiguity study is not warranted. Further research will be required to understand exactly what factors influenced their results, but the simplest explanation is that participants were confused, leading to errors on some portion of the trials. If this same confusion affected the first-mention trials, that may have depressed evidence of first-mention – which in general is not as robust as the gender effect, even in adults – to below detectable levels in the child data.

In sum, there seems little reason to doubt that by five years of age – and, most likely, by two and a half – children are sensitive to first-mention.

TABLE 5. *Comparing five-year-old results for gender trials when the target was the first-mentioned character or the second-mentioned character. Significant time windows are in bold.*

	150 ms	250 ms	350 ms	450 ms	550 ms	650 ms	750 ms	850 ms	950 ms	1050 ms	1150 ms	1250 ms	1350 ms	1450 ms	1550 ms
Target=first-mention															
z	0·75	1·34	**2·24**	1·41	**1·97**	**3·22**	**2·95**	**3·24**	**2·54**	**2·66**	**3·47**	**2·99**	**2·80**	1·79	**2·45**
p	·45	·18	**·03**	·16	**·049**	**·001**	**·003**	**·001**	**·01**	**·01**	**·001**	**·003**	**·005**	·07	**·01**
Target=second-mention															
z	1·93	**2·22**	1·87	**2·62**	**3·65**	**4·37**	**4·82**	**5·27**	**5·65**	**5·18**	**5·61**	**4·81**	**3·54**	**3·74**	**3·19**
p	·054	**·03**	·06	**·01**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**
Difference															
z	1·13	0·41	0·31	0·69	1·22	0·90	1·27	1·66	**2·53**	1·77	1·70	1·44	0·62	1·39	0·43
p	·26	·69	·76	·49	·22	·37	·20	·10	**·01**	·08	·09	·15	·53	·16	·67

TABLE 6. *Comparison of adult results for gender trials when the target was the first-mentioned character or the second-mentioned character. Significant time windows are in bold.*

|	150 ms	250 ms	350 ms	450 ms	550 ms	650 ms	750 ms	850 ms	950 ms	1050 ms	1150 ms	1250 ms	1350 ms	1450 ms	1550 ms
Target=first-mention															
z	0·51	1·64	1·00	0·05	0·50	0·78	1·49	**2·51**	**3·24**	**3·61**	**4·27**	**4·21**	**2·62**	**2·67**	**2·80**
p	·61	·10	·32	·96	·62	·44	·14	**·01**	**·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**	**<·001**
Target=second-mention															
z	0·00	0·93	0·74	0·77	1·63	1·67	1·37	**2·51**	**2·16**	**2·20**	**2·64**	**3·25**	**3·56**	**3·65**	**3·03**
p	1·00	·35	·46	·44	·10	·09	·17	**·01**	**·03**	**·03**	**·008**	**·001**	**<·001**	**<·001**	**<·001**
Difference															
z	0·09	0·11	0·19	0·50	0·87	0·70	0·12	0·03	0·82	1·11	1·29	0·70	0·79	0·97	0·15
p	·93	·91	·85	·62	·38	·48	·90	·98	·41	·27	·20	·48	·43	·33	·88
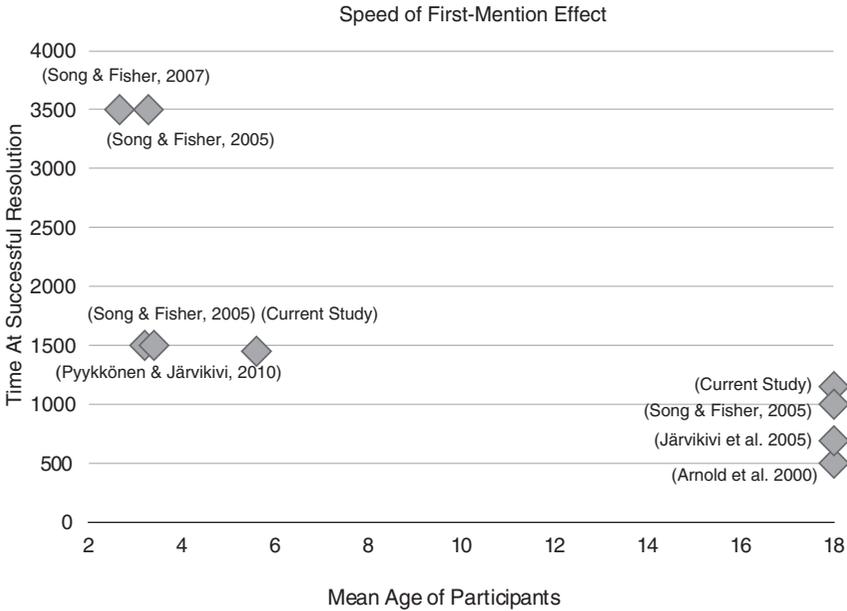
Fig. 7. Plot of results from present and previous first-mention eye-tracking experiments. Mean age of participants is plotted against the earliest evidence of use of first-mention information (the mean time for the first significant window). Adults in Song and Fisher (2005) showed an effect which was strictly significant in the 1000–2000 ms window but significant by items and not subjects in the 0–1000 ms window; the difference was split, and time-of-resolution is shown at 1000 ms. Arnold *et al.* (2007) is excluded because no evidence of use of first-mention information was found.

Several questions remain, however. Inspection of the results across studies that employed participants of different ages suggests that speed of application of first-mention information increases through development (Figure 7). While the results for the very youngest children may bear replication, this trend is potentially an intriguing finding, since – as noted by Arnold and colleagues (2007) – children are not necessarily slower than adults at some other aspects of pronoun processing (see Figures 4–6 and surrounding discussion, above). In the remainder of this section, we consider possible explanations for this finding, should it prove robust.

Account for the developmental speed-of-processing trend

Adults are faster than children at using first-mention information. We consider two – not necessarily mutually exclusive – possible explanations below.

Most-recent mention bias. Children's slow use of first-mention information may be explained by a difficulty in suppressing recent information in favor of

earlier information. The first piece of evidence comes from the fact that children resolved pronouns in the GENDER CONDITION much more quickly when the target referent was the most-recently mentioned character than when it was the first-mentioned character, a difference which was not observed in adults. This result is particularly striking because of the large baseline differences: when the target was the second-mentioned character, five-year-olds were already looking at the target at pronoun onset on 60·5% of the trials, whereas when the target was the first-mentioned character, they were looking at the target on only 40·9% of trials. Thus, although shift of attention to the target would be harder to detect when the target was the second-mentioned character, evidence of such shifts was seen first in that condition. This data suggest that the most recent character may be more activated for children than adults, causing the first-mention bias for ambiguous pronouns to emerge more slowly as it fights against the tide.

Interestingly, recent work has suggested a similar explanation for children's well-known difficulty with syntactic raising. A number of studies (e.g. Choe, 2012; Hirsch & Wexler, 2007) have found that children have difficulty comprehending (6a) but not (6b):

(6)   a. John seems to Mary to be happy.
      b. It seems to Mary that John is happy.

Choe (2012) has suggested that the issue is not syntactic raising per se, but rather due to *Mary* intervening between the predicate *be happy* and its intended argument, *John*. When the word order for (6a) is rearranged so that Mary no longer sits between the predicate and its argument (7), children's difficulty interpreting the sentence disappears:

(7)   To Mary, John seems to be happy.

Thus, it may be that in general children find it easier to attribute new information to the most-recently mentioned entity.

Converging results come from the comparison of the REPEATED-MENTION (SHORT) and REPEATED-MENTION (LONG) conditions. Children resolved the pronoun at least half a second earlier in the SHORT condition than in the LONG condition, and the difference between conditions was significant for every time window from 500 to 1100 ms. In contrast, this effect for adults was much reduced: Although the first strictly significant window was earlier for the SHORT condition (300–400 ms), the LONG condition began trending towards significance at the same time, and at no point was the comparison of conditions significant for adults.

Thus, while analysis of the GENDER condition suggests that children find a very recently mentioned character particularly attractive as a referent for a pronoun, analysis of the REPEATED MENTION conditions suggest that

children have particular difficulty in resolving a pronoun to a character that has not been mentioned recently, even if no other plausible referent has been mentioned in the interim. These two facts in combination should make the FIRST-MENTION condition particularly difficult for children, as they must suppress a recently activated character and re-activate an earlier-mentioned character.

Discourse pragmatics. Recent evidence suggests that recruiting first-mention information to resolve pronouns is different in kind from exploiting gender information. Kehler and colleagues (Kehler, 2002; Kehler *et al.*, 2008) have argued that the first-mention bias is one side effect of our ability to use the structure of the discourse to predict which entity in one sentence will be mentioned again in the next one, and have provided data showing that this bias exists whether or not that re-mention involves a pronoun.

On this account, sentences in a discourse are governed by a discrete set of discourse relations, each of which specifies how the information in a given sentence relates to that in previous sentences. The sentences in our FIRST-MENTION condition were governed by the ELABORATION relation, in which the second sentence elaborates on the situation described in the first (*Michael had lunch with Jacob. He ate a chicken sandwich*). A key feature of this relation is that that the topics of the sentences (typically the subjects) usually co-refer, which results in a first-mention bias if the topic/subject of the second sentence is a pronoun. As Kehler and colleagues note, studies reporting a first-mention bias typically employ sentences governed by the ELABORATION relation or the OCCASION relation, in which the second sentence describes an event that occurred after the first (*John went to the store. He bought ice cream*) and which also strongly prefers co-referring topics.

Other discourse relations work differently (in addition to above, see, *inter alia*, Crinean & Garnham, 2006; Garvey & Caramazza, 1974; Hartshorne & Snedeker, 2013; Stewart, Pickering & Sanford, 1998; Stevenson, Crawley & Kleinman, 1994). For instance, when the RESULT relation holds, the first sentence describes an event and the second sentence describes a result of that event, typically in terms of the consequences to the affected entity. Since the entity affected by an event typically appears as the direct object of the verb (Levin & Rappaport Hovav, 2005), this often results in a second-mention bias:

(8)   Sally frightened Mary so she ...

Importantly, deriving the appropriate pronoun resolution bias depends on inferring the correct discourse relation. Thus, what may be developing is the ability to identify the appropriate discourse relation. The fact that

researchers have not yet uncovered the mechanisms behind this in adults suggests that it may be a complex process requiring time and practice.

Previous studies suggested divergent conclusions about whether children can use first-mention information to resolve pronouns. The present results reconcile these findings. Children at least as young as 2;6 (Song & Fisher, 2007) can use first-mention information. Interestingly, their speed of doing appears to increase during the preschool years (Figure 7). This increase is remarkable and bears further investigation. We do not know when children reach adult levels of proficiency, as the school years remain unstudied. Above, we suggest that this speed-up may be a result of children learning to suppress an incorrect but more recent salient alternative and reactivate earlier information; identifying how children acquire this ability and what changes are necessary to the underlying processes may be a profitable direction for future work. Similarly, these results highlight gaps in our understanding of discourse relations and re-mention biases (cf. Kehler, 2002; Kehler *et al*., 2008): much remains to be learned about when and how children acquire knowledge of discourse structure.

REFERENCES

Arnold, J. E. (1998). *Reference form and discourse patterns*. (Unpublished dissertation), Stanford University.

Arnold, J. E., Brown-Schmidt, S. & Trueswell, J. C. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language & 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 of pronoun resolution from eyetracking. *Cognition* **76**, B13–26.

Bakker, M., van Dijk, A. & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science* **7**, 543–54.

Bates, D. M. & Sarkar, D. (2007). lme4: linear mixed-effects models using S4 classes (Version R package version 0.9975-12).

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

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child Development* **39**, 433–48.

Choe, J. (2012). Acquisition of raising: intervention effects in A-movement. *Proceedings of the 48th Annual Meetings of the Chicago Linguistic Society (CLS)*.

Corbett, A. T. & Chang, F. R. (1983). Pronoun disambiguation: accessing potential antecedents. *Memory & Cognition* **11**, 283–94.

Crawley, R. A. & Stevenson, R. J. (1990). Reference in single sentences and in texts. *Journal of Psycholinguistic Research* **19**, 191–210.

Crawley, R. A., Stevenson, R. J. & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* **19**, 245–64.

Crinean, M. & Garnham, A. (2006). Implicit causality, implicit consequentiality, and semantic roles. *Language and Cognitive Processes* **21**, 636–48.

Ervin, S. M. (1964). Imitation and structural change in children's language. In E. H. Lenneberg (ed.), *New directions in the study of language*. Cambridge, MA: MIT Press.

Fagan, J. F. (1971). Infants' recognition memory for a series of visual stimuli. *Journal of Experimental Child Psychology* **11**, 244–50.

Fantz, R. L. (1958). Pattern vision in young infants. *Psychological Record* **8**, 43–7.

Fernald, A., Perfors, A. & Marchman, V. (2006). Picking up speed in understanding: speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology* **42**, 98–116.

Fernald, A., Pinto, J. P., Swingly, D., Weinberg, A. & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science* **9**, 228–31.

Fernald, A., Thorpe, K. & Marchman, V. A. (2010). Blue car, red car: developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology* **60**, 190–217.

Garvey, C. & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry* **5**, 459–64.

Gordon, P. C. & Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language* **34**, 216–31.

Gordon, P. C., Grosz, B. J. & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science* **17**, 311–47.

Gordon, P. C. & Scearce, K. A. (1995). Pronominaliation and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition* **23**, 313–23.

Hartshorne, J. K. & Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience* **6**(8).

Hartshorne, J. K. & Snedeker, J. (2013). Verb argument structure predicts implicit causality: the advantages of finer-grained semantics. *Language and Cognitive Processes* **28**(10), 1474–1508.

Hirsch, C. & Wexler, K. (2007). The late acquisition of raising: what children seem to think about *seem*. In S. Dubinsky & B. Davies (eds), *New horizons in the analysis of control and raising*, 35–70. New York: Springer.

Jaeger, F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Langauge* **59**, 434–46.

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

Kaiser, E. & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes* **23**, 709–48.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kehler, A., Kertz, L., Rohde, H. & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics* **25**, 1–44.

Levin, B. & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press.

Maratsos, M. P. (1974). Children who get worse at understanding the passive: a replication of Bever. *Journal of Psycholinguistic Research* **3**, 65–74.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, J. T. & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* **57**.

Pyykkönen, P. & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology* **57**(1), 5–16.

Pyykkönen, P., Matthews, D. & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: a visual world study of pronoun resolution. *Language and Cognitive Processes* **25**, 115–29.

R-development-core-team (2005). R: a language and environment for statistical computing. Vienna: R Foundation for Satistical Computing, online: <http://www.R-project.org>.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholiguistic Research* **23**, 197–229.

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

Song, H.-J. & Fisher, C. (2007). Discourse prominence effects on 2·5-year-old children's interpretation of pronouns. *Lingua* **117**, 1959–87.

Spelke, E. S. (1985). Preferential-looking methods as tools for the study of cognition in infancy. In N. A. Krasnegor (ed.), *Measurement of audition and vision in the first year of postnatal life: a methodological overview,*, 323–63. Westport, CT: Ablex Publishing.

Stevenson, R. J., Crawley, R. A. & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes* **9**, 519–48.

Stewart, A. J., Pickering, M. J. & Sanford, A. J. (1998). Implicit consequentiality. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 1031–6. Mahwah, NJ: Lawrence Earlbaum Associates.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* **268**, 1632–4.

Yang, C. L., Gordon, P. C., Hendrick, R. & Hue, C. W. (2003). Constraining the comprehension of pronominal expressions in Chinese. *Cognition* **86**, 283–315.