BUCLD 35 Proceeings
To be published in 2011 by Cascadilla Press
Rights forms signed by all authors

Ambiguous Pronoun Processing Development: Probably not U-Shaped

Joshua K. Hartshorne, Rebecca Nappa, and Jesse Snedeker*

1. Introduction

1.1 The first-mention bias

Unlike proper names, which refer to the same entity every time they are used (1), pronouns have no fixed reference and may refer to different entities every time they are used (2).

(1) a. Jane Austen is an author. **Agatha Christie** wrote many books.
b. Zora Neale Hurston is an author. **Agatha Christie** wrote many books.
c. Ursula LeGuin is an author. **Agatha Christie** wrote many books.

(2) a. Jane Austen is an author. **She** wrote many books.
b. Zora Neale Hurston is an author. **She** wrote many books.
c. Ursula LeGuin is an author. **She** wrote many books.

In many cases, as exemplified by (2), third-person pronouns co-refer with the subject of the previous sentence. Arnold (1998) found that third-person subject pronouns co-referred with the pervious sentence's subject in 64% of cases drawn from a corpus of children's books.

Numerous studies have shown that in many cases adults expect pronouns to co-refer with previous subjects, even in the absence of additional clues to reference and where alternate interpretations are plausible (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Corbett & Chang, 1983; Crawley, Stevenson, & Kleinman, 1990; Jarvikivi, van Gompel, Hyona, & Bertram, 2005). Thus, in (3), most adults prefer that *she* refer to Jane Austen, not Agatha Christie.

(3) Jane Austen$_1$ was born before Agatha Christie$_2$. She$_1$ wrote many books.

* Joshua K. Hartshorne (corresponding author): jharts@wjh.harvard.edu, Department of Psychology, Harvard University. Rebecca Nappa: rnappa@wjh.harvard.edu, Department of Psychology, Harvard University & Department of Brain and Cognitive Sciences, MIT. Jesse Snedeker: snedeker@wjh.harvard.edu, Department of Psychology, Harvard University. The authors wish to thank the members of SnedLab and attendees at BUCLD 35 for comments and suggestions. This work was supported by NDSEG (JKH) and the Ellison Foundation (JS).

It is not clear whether this is properly a bias to have pronouns co-refer with previous subjects or the first-mentioned entity in the previous sentence, or both (Crawley, et al., 1990; Jarvikivi, et al., 2005). As subject-hood and first-mention-ness is heavily confounded in English, we adopt the more common term (*first-mention bias*) without making any strong theoretical commitments.

A number of studies have investigated the development of the first-mention bias in children. Intriguingly, while several studies of children under four years of age indicate adult-like first-mention biases (Pyykkonen, Matthews, & Jarvikivi, 2010; Song & Fisher, 2005, 2007), the one study to look at older children (4-6 years old) found that children failed to show a first-mention bias (Arnold, Brown-Schmidt, & Trueswell, 2007).

1.2 Acquisition of the first-mention bias: The u-shaped development hypothesis

One possible explanation is that the first-mention bias is subject to u-shaped development. U-shaped development is not unprecedented in language acquisition. For instance, children initially make relatively few errors on irregularly-inflected past tense forms (*went, slept*), but the error rate actually increases for 3-4 year-olds before eventually returning to ceiling in adulthood (Figure 1; Cazden, 1968; Ervin, 1974; Marcus, et al., 1992). Similarly, after some initial success, children become less accurate in interpreting passive sentences before eventually recovering (Figure 2; Bever, 1970; Maratsos, 1974).

The traditional account for these u-shaped developmental trajectories is that children initially learn irregular past tense verbs or passive sentences one-by-one in an essentially item-based fashion. Then, at some point, children discover a pattern elsewhere in the language that they over-generalize to exceptional cases. In the case of the past tense, they discover that most verbs form their past tense by adding *–ed*, and they then over-apply this to irregulars. In the case of passives, they discover the dominant active sentence pattern in which the first verbal argument is usually the agent and this causes difficulties until they have fully mastered the exceptional passive pattern.

U-shaped development for the first-mention bias could be explained in a related – if not identical – fashion. By three years of age, children have discovered that in the majority of cases pronouns 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, there are other less common but highly reliable biases that must be acquired as well. In the following sentences, adults reliably expect the pronoun to refer to the second-mentioned character (Garvey & Caramazza, 1974; Hartshorne & Snedeker, under review):

(4) a. John$_1$ liked Bill$_2$ because he$_2$ was from California.
b. Alfred$_1$ feared Theodore$_2$ because he$_2$ was from Kansas.

It may be that when children discover these less frequent patterns, they are initially unsure which types of sentences are governed by which patterns – as, indeed, theorists are unsure – and this caused the older children in Arnold et al. (2007) to make a larger number of errors.

1.3 Acquisition of the first-mention bias: The processing speed hypothesis

An alternative explanation is suggested by the nature of the studies themselves. The studies of children's acquisition of the first-mention bias are preferential-looking/eyetracking studies. For instance, in Arnold et al. (2007), 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.

The dependent measure was whether participants were looking at Donald or at Mickey from the onset of the pronoun (*he's*) to the disambiguating word (*umbrella* – in the illustration, Donald carries an umbrella and Mickey does not). A first-mention effect would be demonstrated by more looks to Donald than to Mickey during the ambiguous region.

With one exception (discussed below), all these studies similarly measured eye movements during an ambiguous region. Table 1 shows the length of the ambiguous regions employed in each study as well as the first time window in which a significant first-mention effect was found. For comparison, two related adult studies are included.

Table 1

Study	Age	Ambig. Region Length	1st Sig. Window
Song & Fisher, 2007	2;6	3700 ms	3000-4000 ms
Song & Fisher, 2005	3;0	3700 ms	3000-4000 ms
Song & Fisher, 2005	3;0	3700 ms	1000-2000 ms
Pyykkonen & Jarvikivi, 2010	3;5	>4000 ms	1240-1760 ms
Arnold et al., 2007	4;0-5;9	650 ms	NA
Arnold et al., 2000	adult	650 ms	400-600 ms
Jarvikivi et al., 2005	adult	>1000 ms	480-690 ms[1]

Preferential-looking/eyetracking studies of the first-mention bias in children. Note that Song & Fisher (2005) contained two relevant experiments.

[1] Jarvikivi et al. (2005) was conducted in Finnish and crossed effect of subject-hood and first-mention-ness. The effect of first-mention-ness independent of subject appeared 690-900 ms after pronoun onset.

Note that Arnold et al. (2007) uses a considerably shorter ambiguous region than any of the other developmental studies. It may be that (a) children are slower at using the first-mention bias, and thus (b) the children in Arnold et al. (2007) simply did not have enough time. The authors reject this possibility because in related pronoun tasks, children are just as fast as adults: specifically, when the pronoun is disambiguated by gender (6).

(6) Donald is bringing some mail to Minnie, while a big rainstorm is beginning. S/he's carrying an umbrella, and it looks like they're both going to need it.

Nonetheless, it is clear from Table 1 that were the other developmental studies restricted to 650 ms time windows, none of them would have shown significant effects. Without testing older children with longer ambiguous regions, this *processing speed hypothesis* cannot be disentangled from the *u-shaped development hypothesis* – or the possibility that some subset of the previous findings are otherwise artifactual and/or do not replicate.

It must be noted that Arnold et al. (2007) do include an additional experiment which is not subject to the same limitations:

(7) Puppy [male] is having lunch with Froggy [female]. He wants some milk.

Participants listened to (7) and then indicated which stuffed animal wanted milk. Children (3;5-4;0 and 4;1-5;0) chose at chance while adults picked the first-mention character (Puppy) nearly 90% of the time.

While this task does remove the issue of the ambiguous time window, these results are difficult to interpret for the following reason: all three participant groups frequently chose the wrong animal even in unambiguous trials (adults did so over 20% of the time):

(8) Puppy [male] is having lunch with Froggy [female]. She wants some milk.

These unexplained findings suggest participants did not fully understand the task, limiting the conclusions that can be drawn.

2. The Experiment

2.1 Overview

In order to contrast the *u-shaped development* and *processing speed* hypotheses, we tested 5 year-olds using paradigm combined the features of the two Arnold et al. (2007) studies. Participants listened to stories while viewing illustrations as their eye movements were tracked. An example image is shown in Figure 1. Examples of stories from each of four conditions are shown in Table 2.

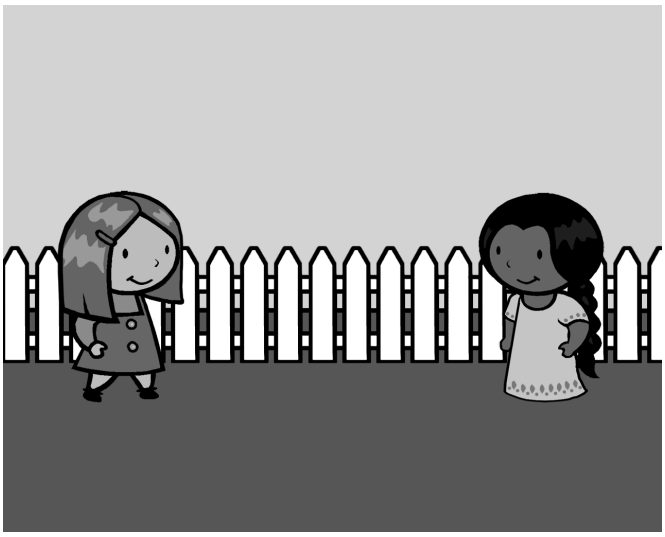

Figure 1. An example image from the experiment. Emily is on the left and Hannah is on the right.

Table 2

condition	example story
first-mention	Emily went to school with Hannah. She read ten books. Can you point to her?
first-mention + repeated-mention (short)	Emily and Hannah are going to Disneyland. Emily has never been to Disneyland. She is really excited about going to Disneyland. Can you point to her?
first-mention + 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. She is really excited about going to Disneyland. Can you point to her?
gender	Emily played baseball with Michael. S/he hit five homeruns. Can you point to her?

Example stimuli from the present experiment.

Note that unlike in the previous eyetracking studies, sentences did not contain disambiguating words, and thus there is no limit on the time in which participants have to respond. In addition, participants provided an off-line pointing response, indicating explicitly who they thought the pronoun refers to.

The core condition of interest is the *first-mention condition*. If the *u-shaped development hypothesis* is correct, 5 year-olds should fail to look at the first-mentioned character or point at the first-mentioned character at above-chance rates. If the *processing speed hypothesis* is correct, they should do both; however, looks to the first-mentioned character should not rise above chance until more than 650 ms after pronoun onset.

The *first-mention + repeated-mention (short)* and *first-mention + repeated-mention (long)* conditions broaden the discussion by providing participants with two types of information: one character is mentioned first and mentioned again, which should serve to make that character particularly salient. The difference between conditions is that the *long* version has two additional filler sentences between the final-mention of the target character and the pronoun, sentences which mention neither character.

Finally, in the *gender condition*, the gender of the pronoun eliminates one of the characters from consideration. As noted in the example sentence (Table 2), the character with matching gender was equally likely to be the first-mention or last-mention character.

2.2 Method

2.2.1 Participants

Twenty-two five year-olds (5;0-5;11; M=5;7) and 25 adults participated. One additional adult was excluded for excessive track loss. All participants were native English speakers.

2.2.2 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 midline. Characters were not engaged in any activities, and the background was neutral (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. These stories were created and counter-balanced as follows.

Four *first-mention condition* stories were created (Table 2), two with male characters and two with female characters. For any given participant, every character appeared in first-mention position. For any given story, which participant was first-mention was counter-balanced across participants; thus there was a variant of the story in Table 2 in which Hannah went to school with Emily.

Eight *first-mention + repeated-mention condition* stories were created, each with two optional filler sentences that mentioned neither character (see Table 2). Any given participant saw four of these scenarios with the filler sentences (the *long* condition) and four without (the *short* condition). Each story appeared in both conditions, counterbalanced across participants. As in the *first-mention*

condition, for any given participant, each of the characters appeared in first-mention position in the *first-mention + repeated-mention (short)* and in the *first-mention + repeated-mention (long)* conditions, and for any given story, which character was mentioned first was counter-balanced across participants.

Finally, eight *gender condition* stories were created, such that each character was first-mentioned twice – once with each possible opposite-gender character – once as the target (with a pronoun of congruent gender) and once not (the pronoun's gender matched the other character). For any given story, the pronoun (*he/she*) was counter-balanced across participants. For some participants, one sound file had the wrong-gendered pronoun in the pointing instruction (*Can you point to her*), and thus for those participants eye movements after that final object pronoun were discarded and the pointing response was not counted.

The counter-balancing 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 was used for all four lists. Four additional lists were created by reversing the order of the stimuli, for a total of eight lists.

2.2.3 Procedure

Participants where familiarized with the characters and drilled on their names until they could name each character rapidly and without hesitation. Stimuli were presented on a Tobii T60 desktop eyetracker, which also recorded eye movements. Participants' offline responses (pointing) were recorded by the experimenter. An approximately equal number of participants were tested with each stimulus list.

2.3 Results

2.3.1 Offline (pointing) results

Participants' offline responses are shown in Table 3. In all conditions, both five year-olds and adults were more likely to choose the target character than would be expected by chance alone (*p*s<.05). In fact, adults were at ceiling in all conditions, and five year-olds were at ceiling in the *gender condition*.

Table 3

Condition	% Choosing Target
Adults	
First-Mention	97% (2%)
First-Mention + Repeated-Mention (Short)	96% (2%)
First-Mention + Repeated-Mention (Long)	100% (0%)
Gender	100% (0%)
Five Year-Olds	
First-Mention	70% (5%)
First-Mention + Repeated-Mention (Short)	85% (4%)

First-Mention + Repeated-Mention (Long)	85% (4%)
Gender	99% (1%)

Offline (pointing) results with standard errors.

2.3.2 Online (eyetracking) results

At pronoun onset, participants were usually looking at either the target character or the distractor character. Thus, we analyzed shifts off of the initially-fixated character. We coded each trial in terms of whether the participant was initially fixating the target or distractor. During each 200 ms bin from pronoun onset, we coded whether the participant was still fixating the same character or was now fixating the other character. The critical question was at what point the proportion of participants who initially fixated the distractor and were now fixating the target – that is, those who switched in the "correct" direction – was greater than the proportion of participants who initially fixated the target and were now fixating the distractor – that is, those who switched in the "wrong" direction.

Results were analyzed using mixed-effects models with subjects and sentence templates as random effects in R (R-development-core-team, 2005) using the lme4 package (Bates & Sarkar, 2007), and p-values were estimated using the function pvals.fnc, which implements Markov chain Monte Carlo sampling with 10,000 samples (see Baayen, 2008, for discussion).

More adults had switched to target by 1200-1400 ms in the *first-mention condition* (t=2.12, p=.03), by 400-600 ms in the *first-mention + repeated-mention (short) condition* (t=2.18, p=.03), by 800-1000 ms in the *first-mention + repeated-mention (long) condition* (t=3.15,p=.003), and by 600-800 ms in the *gender condition* (t=2.16, p=.03). More five year-olds had switched to target by 1400-1600 ms in the *first-mention condition* (t=2.87, p=.008), by 200-400 ms in the *first-mention + repeated-mention (short) condition* (t=2,22, p=.03), by 1200-1400ms in the *first-mention + repeated-mention (long) condition* (t=2.99, p=.004), and by 400-600 ms in the *gender condition* (t=2.46, p=.02).

3. General Discussion

Five year-olds preferred resolving pronouns to the first-mention/subject character in the previous sentence, an effect which appeared in both online eye movements and ultimate interpretation. The eye movement effect did not emerge until 1200-1400 ms after pronoun onset, consistent with Pyykkonen & Jarvikivi's (2010) results with children aged 3;5. Importantly, the ambiguous region used by Arnold et al. (2007) was too brief for such an effect to have emerged. Thus, the present results support the claim that children are able to employ first-mention information more rapidly as they develop (the *processing speed hypothesis*). We return to this notion after discussing the remaining data.

When one character was both first-mentioned and repeatedly mentioned, both adults and children resolved subsequent pronouns to that character, and

both adults and children were slower to do so when two sentences passed between the most recent mention and the pronoun. However, this effect was much larger for the five year-olds, who were both faster in the *short* condition and slower in the *long* condition relative to adults. These results may suggest that children are over-influenced by the salience of the most recently-mentioned discourse entity and have difficulty recovering entities mentioned earlier.

The puzzle posed by Arnold et al. (2007) was the following: why do children show adult-like processing of pronoun gender information but not first-mention information? It is now clear that children even in the age range studied by Arnold et al. do in fact use first-mention information, but it remains the case that they do not do so in an adult-like fashion: that is, they are much slower than adults to use first-mention information, and this is not the case for gender information. Thus, the question remains.

We consider several possible answers. First, Arnold et al. (2007) suggested that the efficient use of first-mention information was acquired later than the use of gender because the former is a less reliable cue to pronoun resolution than the latter. It may be that even once children notice less reliable cues, it takes additional time and practice for them to be able to use them rapidly.

Second, it may be the case that using first-mention information to constrain utterance interpretation involves top-down processing. Based on studies of NP-attachment, Snedeker, Worek and Shafto (2010) argue that an important factor in language development is increasing facility with using top-down constraints.

A third possibility is that children and adults arrive at the first-mention bias through fundamentally different processes. Some authors argue that pronoun resolution is the result of complex inferences based in part on suppositions about the nature of the discourse and the speaker's intentions (Kehler, Kertz, Rohde, & Elman, 2008), while others describe the same phenomena in terms of simple heuristics (Arnold, et al., 2007; Arnold, et al., 2000; Van Berkum, Brown, Zwisterlood, Kooijman, & Hagoort, 2005). It may be that the youngest children have not yet developed heuristics and thus must engage in complex inferential processing, whereas during development they develop simple heuristics that speed processing.

One additional possibility is that children's difficult consists of re-activating previously-mentioned discourse entities. This is consistent with their relative difficulty in both the *first-mention* and the *first-mention + repeated-mention (long)* conditions.

Current data do not allow us to choose between these options. However, it is clear that the determining the right account will have broad implications for language acquisition and linguistic processing theory in general.

References

Arnold, Jennifer E. (1998). *Reference form and discourse patterns.* Dissertation, Stanford University.

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

Arnold, Jennifer E., Eisenband, Janet G., Brown-Schmidt, Sarah, & Trueswell, John C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76*, B13-B26.

Baayen, R. H. (2008). *Analyzing Linguistic Data*. Cambridge, UK: Cambridge University Press.

Bates, Douglas M., & Sarkar, Deepayan (2007).

lme4: Linear mixed-effects models using S4 classes (Version R package version 0.9975-12).

Bever, Thomas G. (1970). The cognitive basis for linguistic structures. In John R. Hayes & Roger Brown (Ed.), *Cognition and the Development of Language*. New York, NY: Wiley.

Cazden, Courtney B. (1968). The acquisition of noun and verb inflections. *Child Development, 39*, 433-448.

Corbett, Albert T., & Chang, Frederick R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition, 11*, 283-294.

Crawley, Rosalind A., Stevenson, Rosemary J., & Kleinman, David (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, 19*, 245-264.

Ervin, Susan M. (1974). Imitation and structural change in children's language. In E. H. Lenneberg (Ed.), *New Directions in the Study of Language*. Cambridge, MA: The MIT Press.

Garvey, Catherine, & Caramazza, Alfonso (1974). Implicit causality in verbs. *Linguistic Inquiry, 5*, 459-464.

Hartshorne, Joshua K., & Snedeker, Jesse (under review). What is implicit causality: world knowledge, an arbitrary feature or an effect of semantic structure.

Jarvikivi, Juhani, van Gompel, Roger, Hyona, Jukka, & Bertram, Raymond (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject preference accounts. *Psychological Science, 16*, 260-264.

Kehler, Andrew, Kertz, Laura, Rohde, Hannah, & Elman, Jeff L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*, 1-44.

Maratsos, Michael P. (1974). Children who get wordse at understanding the passive: A replication of Bever. *Journal of Psycholinguistic Research, 3*, 65-74.

Marcus, Gary, Pinker, Steven, Ullman, Michael, Hollander, Michelle, Rosen, T. John, Xu, Fei, et al. (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development, 57*.

Pyykkonen, Pirita, Matthews, Danielle, & Jarvikivi, Juhani (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*, 115-129.

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

Snedeker, Jesse, Worek, Amanda, & Shafto, Carissa L. (2010). *The role of lexical bias and global plausibility in children's online parsing: A developmental shift from bottom-up to top-down cues.* Paper presented at the 34th Annual Boston University Conference on Language Development.

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

Song, Hyun-joo, & Fisher, Cynthia (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua, 117*, 1959-1987.

Van Berkum, Jos J. A., Brown, Colin M., Zwisterlood, Pienie, Kooijman, Valesca, & Hagoort, Peter (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443-467.