Ψ **Psychology Press**
Taylor & Francis Group

Children's use of gender and order-of-mention during pronoun comprehension

Jennifer E. Arnold

University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Sarah Brown-Schmidt

University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA

John Trueswell

University of Pennsylvania, Philadelphia, PA, USA

Two experiments were conducted to examine the on-line processing mechanisms used by young children to comprehend pronouns. The work focuses on their use of two highly relevant sources of information: (1) the gender and number features carried by English pronouns, and (2) the differing accessibility of discourse entities, as influenced by order-of-mention in a clause. Adults use both evidential sources, as early as 200 ms after the offset of the pronoun (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000). We find that like adults, 3–5-year-old children use a pronoun's gender to guide their choice of a referent, and that they use it rapidly on-line. But unlike adults, they show little or no signs of a first-mentioned bias, either off-line or on-line. This is consistent

---

---

with a tendency for children to initially recruit reliable sources of constraint for language comprehension – in this case, the gender of the pronoun.

INTRODUCTION

Human languages each possess a range of ways of referring to entities in the world. Noun phrases can be complex, and rich in their semantic content (e.g., *the girl you met yesterday*) or simple, and blanched of almost all of their meaning (*she*). Somewhat surprisingly, native speakers of a language rarely have any difficulty deciding on the degree of specificity needed in their expressions, and listeners in turn find it relatively easy to interpret these expressions, even less specified ones, like pronouns. The speed at which reference is achieved is also impressive; listeners can identify the referent of a full noun phrase, or even a pronoun, within 200–400 ms after the *onset* of the referring expression (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000a; Dahan, Tanenhaus, & Chambers, 2002). Clearly, listeners are quite skilled at reference resolution and presumably have developed highly specialised strategies for mapping linguistic forms onto a mental model of the discourse and the world.

In this paper, we focus on understanding the skill of reference resolution, and how children acquire it.[1] We want to know both what children know about their language, and also how they recruit information from different sources during the moment-by-moment process of interpreting linguistic input as it occurs. Of particular interest will be understanding how reference resolution is constrained by two different sources of information: (1) the lexical content of the referring expression itself; and (2) the properties of the recent discourse, which are known to shape the mental 'accessibility' of possible referents. As a case study, we will focus on the comprehension of pronouns like 'he' and 'she'. In contrast to full noun phrases, pronouns carry relatively impoverished lexical information, typically only about the animacy, gender, and number of a referent. As a result, pronominal referential expressions often force listeners to rely more heavily on the accessibility of possible referents to achieve reference resolution. Thus pronouns are particularly interesting from a processing and acquisition perspective, because they allow us to observe how the features of a lexical item are integrated with information about discourse and referent accessibility.

In the remainder of this introduction, we will focus on the processes of understanding pronominal expressions. We begin by discussing what is known about how adults determine the referents of pronouns, and then we turn to the child's ability and theories of how adult performance is achieved.

---

[1] Preliminary analyses of the data reported here appeared in Arnold et al. (2001) and Arnold, Brown-Schmidt, Trueswell, & Fagnano (2004).

Pronominal reference resolution by adult listeners

Adults are known to use at least two distinct sources of information when computing the most likely referent of a pronoun: (1) the overlap between the pronoun's lexical features (gender, animacy, and number) and the meaningful properties of possible referents; and (2) the accessibility of possible referents, such that the pronoun is linked with the most accessible entity that matches the lexical features (Arnold, 1998; Garnham, 2001; Givón, 1983; Gundel, Hedberg, & Zacharski, 1993; Sanford & Garrod, 1981; Stevenson, Crawley, & Kleinman, 1994).[2]

Of these two sources of information, the relative accessibility of possible referents is likely to be the more complex one to compute. Even for adults, it would be resource-intensive to maintain a detailed model of the knowledge, goals, and intentions of their conversational partners, and compare that with their own knowledge – all of which are potentially relevant for computing common ground and accessibility. One possible heuristic for adults to estimate the common ground is by paying attention to what is linguistically or visually co-present (Clark & Marshall, 1981). The linguistic context is particularly useful in directing listeners to precisely those referents that are most important to the discourse. Indeed, adults do make use of a variety of textual factors for the purposes of reference resolution, including the recency with which potential referents have been mentioned, the grammatical functions and thematic roles of recently mentioned referents, parallelism of the syntactic roles of anaphors and antecedent, and information carried by focus constructions, like clefts (Arnold, 2001; Arnold et al., 2000a; Ariel, 1990; Gernsbacher, 1989; Gernsbacher & Hargreaves, 1988; Givón, 1983; Gordon, Grosz, & Gilliom, 1993; Grosz, Joshi, & Weinstein, 1995; Gundel et al., Sanford & Garrod, 1981; Stevenson et al., 1994; see Arnold, 1998, for a review).

In this paper we will especially focus on the first-mention bias, which is the tendency for adults to consider the first out of two or more entities in a sentence as the more salient one. Often the first-mentioned is also the grammatical subject, which also tends to be considered highly accessible (see Kaiser & Trueswell, in press). This entity thus tends to be the most accessible for the purpose of interpretation of pronouns (Arnold et al., 2000a; Gordon et al., 1993) and fluent deaccented definite noun phrases (Dahan et al., 2002).

Both order-of-mention and gender have been shown to affect immediate, on-line pronoun processing in adults. Pronouns are generally used for

---

[2] Syntactic constraints also exist, especially with regard to pronouns and any intra-sentential antecedents (e.g., Thornton & Wexler, 1999). The constraints are not discussed here because all antecedents in the present experiments were extra-sentential.

referring to things that are highly accessible in the discourse context, and first-mentioned entities are generally taken to be more accessible than other entities (Brennan, 1995; Brennan, Friedman, & Pollard, 1987; Gernsbacher & Hargreaves, 1988; Stevenson et al., 1994). Similarly, comprehenders tend to find pronouns easier to understand when they refer to first-mentioned, rather than later-mentioned entities (Gordon, Grosz, & Gilliom, 1993; McDonald & MacWhinney, 1995, etc.). Not surprisingly, adults also tend to map 'he' onto male characters, and 'she' onto female ones. Even though some researchers have suggested that gender information is secondary to discourse-based salience (Garnham, Oakhill, & Cruttenden, 1992; Greene, McKoon, & Ratcliff, 1992; McDonald & MacWhinney, 1995), others have found that gender influences the earliest moments of pronoun comprehension (e.g., Boland, Acker, & Wagner, 1998, MacDonald & MacWhinney, 1990).

For instance, the rapid effects of both gender and order-of-mention information on pronoun comprehension have been demonstrated by Arnold et al. (2000a), using a spoken-language comprehension task with a visual context. Participants' eye movements were monitored while they looked at pictures with two cartoon characters (see Fig. 7), and listened to stories like 'Donald is bringing some mail to {Mickey/Minnie}, while a violent storm is beginning. {He's/She's} carrying an umbrella, and it looks like they're both going to need it.' Arnold et al. manipulated whether the pronoun referred to the first- or second-mentioned character from the first sentence, and whether the two characters had the same or different gender, which made the pronoun either ambiguous or unambiguous by gender. Eye movements revealed rapid looks to the target character when the pronoun was unambiguous by gender (whether it referred to the first or second character), and when an ambiguous pronoun referred to the first-mentioned character (within 200 ms after the pronoun offset in all cases). Participants did not converge on the target when the pronoun was ambiguous and referred to the second-mentioned character. These results established that both gender and order-of-mention affect the early moments of pronoun comprehension.

One proposal for how adults compute accessibility so rapidly is the expectancy hypothesis (Arnold, 1998, 2001; Arnold & Tanenhaus, in press). Under this view, adults use a range of linguistic and nonlinguistic evidence to estimate what a speaker might refer to in the immediately upcoming utterance. This view builds on the assumption that linguistic features of the discourse history, like order-of-mention, are relevant to accessibility computations because they are linked to the cognitive state of the speaker and listeners, which can also be influenced by nontextual factors (Ariel, 1990; Chafe, 1976, 1994; Gundel et al., 1993). In particular, the text provides important indicators of accessibility because it reveals what the speaker is focusing on. Linguistic co-presence is a strong signal that something is

known to all discourse participants, and the details of how something has been dealt with in the discourse is informative about the degree to which it is prominent in the speaker's mind. Thus, the linguistic history provides the building blocks for coordinating a detailed model of the common ground, and which entities are more accessible than others (Arnold, 1998; Arnold & Tanenhaus, in press).

In particular, the expectancy hypothesis proposes that features like order-of-mention are informative because they correlate with the likelihood that the speaker will mention something again in the immediately upcoming discourse – the best evidence that a speaker considers that thing to be important. If the speaker says 'Elmo called Cookie Monster', as opposed to 'Cookie Monster received a call from Elmo', the listener knows that Elmo is likely to be more central to the discourse than Cookie Monster. Indeed, first-mentioned/subject entities are more likely to be mentioned in the subsequent utterance than other entities (Arnold, 1998, 2001; Arnold & Tanenhaus, in press). This and other features of the discourse lead some entities to have high 'expectancy', and when the listener focuses on those entities, it facilitates comprehension. In this sense, discourse information indicates accessibility not just by looking 'backward' to what has been accessible before, but by looking 'forward' to what is likely to continue to be important to the speaker (Givón, 1983; Grosz et al., 1995).

The expectancy hypothesis contrasts with the traditional approach to explaining accessibility, which focuses on the linguistic features of the discourse history, as described above. The traditional approach requires a separate explanation for each linguistic feature and its relationship to accessibility, whereas the expectancy hypothesis provides an explanation for why such a diverse set of features have the same effect on the use and understanding of pronominal references. It further makes predictions that reference comprehension should be influenced by any factor that modulates the expectancy of a referent, linguistic or nonlinguistic.

Note that there are at least two ways in which expectancy could have its effect. The expectancy hypothesis is built on the observation that certain types of entities tend to be referred to more than others (Arnold, 1998). Listeners could use these regularities directly to calculate the likelihood that a referent will be mentioned. As such a rather unsophisticated statistical mechanism can be at work for anticipating reference. But the reason that language is structured this way is because speakers produce language systematically, due to their own goals, intentions, and the constraints of the planning and production system itself.[3] For example, speakers tend to

---

[3] Similarly, MacDonald (1999; Race & MacDonald, 2003; Gennari & MacDonald, in press) has proposed that constraints on production lead to systematic lexical and syntactic distributions, which comprehenders make use of.

produce accessible information early in an utterance (Arnold, Wasow, Losongco, & Ginstrom, 2000b; Bock, 1982, 1986; Bock & Irwin, 1980; Ferreira, 2003; Ferreira & Dell, 2000). Therefore expectancy information could also be used to make inferences about what is accessible to the speaker. In this way, one could bootstrap into a more sophisticated model-approach, in which the listener begins to develop a 'model' of speaker's intentions and goals and uses this to generate reference expectation. For current purposes it is not possible or necessary to distinguish these possibilities.

Pronominal reference resolution by child listeners

Given the discussion above, it seems reasonable to assume that for children to achieve successful pronoun resolution, they must understand that: (1) the lexical content of referential expressions tends to match the semantics of their referents; (2) accessibility of entities tends to also determine reference (e.g., for pronouns, the most accessible entity is the most likely referent).

The first of these two constraints is true of essentially all referential expressions, not just pronouns, thus one might expect that children would adopt this constraint quite early. Indeed, from a very young age, children learn that words refer to things that match the lexical input to a certain degree.[4] That is, 'dog' refers to a type of animal, 'ball' refers to some round object, etc. Thus, since lexical information tends to be very reliable, it is predicted to play a central role in children's initial, on-line hypotheses about what a pronoun refers to.

But what about accessibility? Under the expectancy hypothesis described above for adults, children need to learn how to compute what is accessible in the discourse. The evidence for this comes from regularities in the way people talk, e.g., they tend to talk more about entities that they have previously placed in first-mentioned position than other entities.

According to this view, the acquisition of any particular cue to accessibility, like order-of-mention, requires experience with the language to amass a database from which the learner can extract the relevant distributional information. This would predict that the strength of any particular cue would be linked to its availability and reliability in the input. For example, if a child observes a consistent tendency for speakers to talk about the things they were just talking about (i.e., a recency effect), the child will learn that recently mentioned things are important to speakers, and thus should be considered accessible. A subtler cue, like order-of-mention, which

---

[4] Of course, this is a reflection of the primary goal of word learning, which is to develop a semantic representation of each lexical item. Our only point here is that there should be an interpretable relationship between what children think the word means and what it is referring to, even if the relationship is novel in some situations (cf. Clark, 1983).

picks out a particular recently mentioned character, may take longer to learn. However, accessibility can be assessed on the basis of many sources of information. Real-life situations typically provide rich physical and linguistic context, and children may pool several partially acquired sources of information to estimate an entity's accessibility in the joint discourse model.

The expectancy hypothesis, as one example of a constraint-based model of reference resolution, makes similar predictions to other constraint-based views of acquisition and processing. For example, the 'multiple cues' model (Trueswell & Gleitman, 2004) suggests that children, like adults, are sensitive to multiple constraints on utterance and word meaning, but they do not treat all sources of evidence equally. Much of this work examines syntactic processing in young children, and how children's referential processing abilities impact on-line parsing (e.g., Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Snedeker, Thorpe, & Trueswell, 2001; Trueswell, Sekerina, Hill, & Logrip, 1999). This line of research suggests that children initially pay attention to the most reliable evidence available – that is, evidential sources that are both frequently available and strongly supportive of a single interpretation for a given input (Bates & MacWhinney, 1989). This leads children to initially show parsing preferences based more on reliable lexical biases, such as whether a verb tends to be used with a prepositional phrase argument or not. By contrast, children do not initially make use of the referential context in the same way as adults do (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell et al., 1999). When applied to the process of pronoun resolution, this view would predict that children should initially rely more on gender, which provides a consistent mapping between pronouns and their referents, compared with the probabilistic and less reliable link between order-of-mention and accessibility.

In sum, the on-line comprehension of referring expressions requires children to rapidly assess and integrate both detailed lexical information, and their representation of the current situation, in particular the accessibility of discourse entities. One might expect that evidence from lexical features (such as gender) would play an important early role in constraining referent resolution in children, whereas proficient estimation of referent accessibility may show a slower developmental trajectory.

By contrast, alternative explanations of discourse cues to accessibility focus on the inherent properties of these positions, for example that grammatical subject or first-mention positions are inherently perceived as topical, and worthy of greater attention. Alternatively, one might suggest that the first-mentioned entity in an utterance enjoys natural prominence as a result of primacy (e.g., Song & Fisher, 2005). Either of these positions would predict that the first-mentioned bias for pronoun comprehension should emerge as a strong constraint very early in a child's development, perhaps

before children become sensitive to lexical information, like gender, which is language-specific.

An examination of the literature offers several points in support of the prediction that children acquire the use of lexical information, like pronominal gender, before they become adept at using accessibility information. However, evidence about the on-line use of both gender and order-of-mention is limited. Several studies suggest that preschoolers can use gender for pronoun interpretation, at least for 'off-line' interpretation, when they have plenty of time. For example, Wykes (1981) found a high rate of accuracy when 4 and 5-year-olds acted out sentences like 'Jane found John's ball. She gave it to him', where the pronouns were disambiguated by gender. By comparison, performance was lower in sentences where the two characters had the same gender and an inference was needed for successful pronoun interpretation. Similarly, Brener (1983) found that children ages 2;8 to 5;7 tended to correctly identify the agentive character in sentences like 'She drank the milk', when there was one male and one female present.

By contrast, evidence about whether children use gender during on-line processing is inconclusive. Tyler (1983) provided some of the earliest on-line evidence of referential processing in children, using a mispronunciation-detection method with Dutch-speaking children ages 5, 7, and 10. The mispronunciation occurred on a word following an anaphoric expression; if the expression were fully understood, it would make the word predictable, and therefore the mispronunciation should be easier to detect. She found that 5-year-olds differed from older children and adults, in that they appeared to have more difficulty processing pronouns than more explicit referring expressions, in conditions where the pronoun was lexically unambiguous, but did not refer to a clearly established discourse topic. However, in the experiment that investigated the lexical features of pronouns most directly, 5-year-olds were significantly faster in comprehending pronouns that were unambiguous by gender, compared with those that were unambiguous by number, leading Tyler to conclude that 5-year-olds may already be starting to make use of gender for pronominal interpretation. If this is true, it would be consistent with other evidence that even younger children can map semantic information from full noun phrases to their referents within 600 to 700 ms of the noun onset (e.g., Swingley, Pinto, & Fernald, 1999).

There is also evidence that children can use accessibility information for pronoun resolution, but perhaps not in a completely adult-like manner. For instance, Song and Fisher (2005) report evidence suggesting that 3-year-old children do understand the link between pronoun use and accessibility. They presented children with stories in which a single character was strongly marked as the most accessible one (e.g., through initial mention in two or more consecutive sentences), and found that children looked more at the referent of a pronoun when it was the most accessible character. However,

this preference tended to emerge about 1 second after the pronoun onset, considerably later than in adult studies (Arnold et al., 2000a). Similarly, Tyler (1983) found in a posthoc analysis of her Exp. 3 results that 5-year-olds were better at pronoun interpretation when the pronoun referred to a clearly defined discourse topic that had been established in two or three preceding sentences. In fact, when there is only one potential match for a pronoun in the discourse, children begin to identify the referent around 400 ms after the onset of the pronoun (Arnold, 2006). There is additional evidence that children know that pronouns are for accessible characters coming from their own productions. When children are focusing on a character, they tend to refer back to it with a pronoun. This is supported by evidence that they use pronouns for referring to entities they have just produced as a grammatical subject (Hickmann & Hendriks, 1999), or the discourse topic (Karmiloff-Smith, 1985). Thus, children may understand that pronouns are used for highly accessible referents, at least when accessibility is strongly determined, but it is not clear how rapidly this information affects on-line processing.

However, it is less clear whether children are estimating accessibility of referents in a way similar to adults. The question relevant to this paper is specifically whether children have learned how to use order-of-mention as a guide to referent accessibility. Adults use this information consistently, and take it into account as rapidly as gender information for the purpose of interpreting pronouns. In Song and Fisher's (2005) experiments, order-of-mention was one of the available cues about referent accessibility, and they concluded that it guided children's eye movements. But crucially, children in these studies always had multiple, convergent linguistic cues about who was the most accessible character (e.g., order-of-mention, repeated mention over consecutive clauses, and previous pronominalisation). One experiment did minimise the amount of accessibility information provided to children (Experiment 4). However, this study still used two context sentences, both of which mentioned the target character first (i.e., order-of-mention was repeated in the story prior to hearing the test pronoun). According to the expectancy hypothesis, multiple sources of information can be used in concert, and when they point to the same referent, they provide a stronger constraint on referent accessibility. Thus, repeated mention of a character in a prominent position, like first-mentioned position, provides stronger evidence of this character's accessibility than just a single mention.

In sum, it is not known whether children can use order-of-mention alone to guide pronoun processing, or whether it is a weak cue that 3- and 4-year-olds are only beginning to discover. Evidence about the use of pronominal gender is in general positive, but again, the data are not strongly conclusive, and it is especially unclear whether young children use gender immediately during on-line processing of pronouns. Finally, a systematic

examination of how these two sources of information combine during on-line processing has not been done.

The present experiments specifically investigate the ability of 4- and 5-year-old children to resolve pronouns on the basis of gender information and order-of-mention as an indicator of referent accessibility. We want to know whether children can recruit and integrate these two sources of information on-line, as rapidly as adults do. We investigate the use of these two constraints during pronoun resolution in two experiments. In both experiments we use spoken language in the context of visually concrete environments. These are characteristics of the environments in which children normally use language, and therefore should facilitate children's performance, allowing us to observe what information, if any, guides their pronoun comprehension. In both experiments reported here, children viewed a visual scene that could serve as their memory for the characters, potentially freeing up mental resources to devote to the comprehension tasks (see Ballard, Hayhoe, Pook, & Rao, 1997). This context also served as a visual reminder of the gender of the characters, who were stereotypical male and female characters.

The first experiment examines off-line pronoun interpretation in children ages 3;6 to 5;0, with supporting on-line data from children's eye movements as they performed the off-line task. The second experiment more closely examines the use of gender and order-of-mention in on-line processing in children ages 4;0 to 5;9.

EXPERIMENT 1

Method

Participants.   We report data from 47 children in two age groups: (1) 24 children aged 3.5–4.0 (43–48 months, average 45.3), 12 boys and 12 girls, and (2) 23 children aged 4.1–5.0 (49–60 months, average 54.3 months), 13 boys and 10 girls. Data from an additional 5 children were excluded because they failed to pass the diagnostic trials ($n = 2$), their parent interfered ($n = 2$), or experimenter error ($n = 1$). All children were native speakers of English, and their parents were recruited from a database of well-baby births in Rochester, NY.

Procedure.   Children sat at a table across from two experimenters (see Figure 1), and listened to a puppet (Elmo) tell simple short stories about two characters, e.g., 'Bunny is playing outside with Froggy. She wants a ball'. The stories were always about two of the following four characters: Froggy, Bunny, Panda Bear, and Puppy. The characters were visually represented by dolls on a table in front of the child. Bunny and Froggy had stereotypical
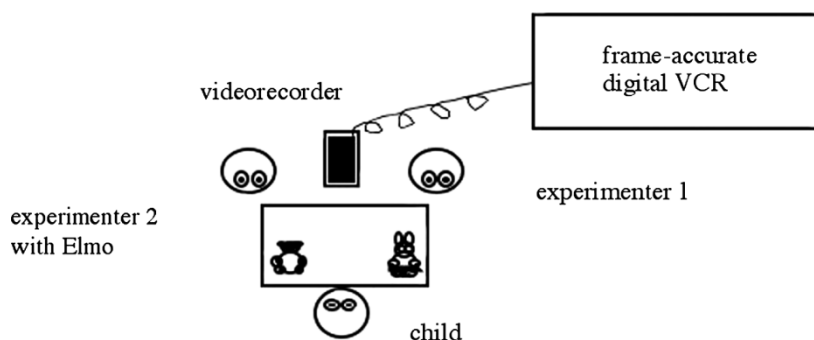
Figure 1.   Experimental setup for Experiment 1.

female appearances, and Puppy and Panda Bear had stereotypical male appearances. During an initial play period we familiarised the children with the four dolls, the two experimenters, and the narrator puppet. The names of the characters were repeated frequently during the play session. At the end of this warm-up session we tested the children's knowledge of the characters' names and genders, by asking them to help the narrator puppet, who purportedly had trouble remembering names and knowing who is a boy and who is a girl. Most children were able to immediately name the characters and identify their genders; if they weren't we practiced until we were confident the child had learned them.

We then asked children to listen to the puppet tell very short stories (two sentences long) and answer a question about each one. Before each story, a standard sequence was followed to introduce the characters and engage the child's attention:

1. Experimenter 1 announced the story, e.g., 'This story is about Froggy and Puppy'. The order of the characters in this statement was counterbalanced within subjects with respect to the order in which they were mentioned in the story.
2. Experimenter 2 placed one of the characters on the table and named it, e.g., 'This is Froggy'. The animal on Experimenter 2's side of the table was counterbalanced between subjects with respect to the order in which it was mentioned in the story.
3. Experimenter 1 placed and named the other character.
4. The child was prepared to hear the story, e.g., 'Are you ready?', or 'Elmo, are you ready to tell another story?' At this point Elmo hides under the table ('Bye bye Elmo'). The child was informed early in the experiment that Elmo's voice would come out of the speakers under the table.

This sequence anchored the beginning of each trial, and served to remind children of the characters' names throughout the experiment.

The first sentence of each story mentioned the two characters doing some reciprocal action, and the second sentence explained that one character wanted a particular item, e.g., 'Puppy is having lunch with Froggy. He wants some milk'. Reciprocal predicates like these are advantageous because they avoid confounds with thematic roles that can sometimes alter the first-mentioned bias (e.g., Garvey & Caramazza, 1974), and they show the typical effect of making the first-mentioned character more accessible for reference with pronouns (Arnold & Griffin, 2006). Experimenter 1 then presented the child with the object, e.g., a toy carton of milk, and asked 'Can you show me who wants the milk in Elmo's story?' The child responded by picking up the toy and placing it in front of one of the two character dolls. During each story, both experimenters looked at the table at a point mid-way between the two characters, so as not to bias the child's response. If the child asked for confirmation (e.g., saying 'Puppy?'), or looked at the experimenters questioningly, the experimenters said 'What did Elmo say?' If the child wasn't sure, the auditory stimulus was played again, until the child made a response.

An initial practice item was used to make sure the children understood the task. This practice item was spoken by experimenter 2. This was followed by 14 pre-recorded stories (9 experimental stimuli and 5 fillers). The experimental items manipulated the gender of the characters (same or different), and in the case of the different-gender conditions, whether the pronoun matched the gender of the first or second character. This resulted in three conditions: different-gender/first-mention; different-gender/second-mention, and same-gender (see Table 1).

The five filler trials were identical in format to the experimental stimuli, except that a name was used in the second sentence, e.g. 'Panda Bear and Puppy are playing in the bathtub. Puppy wants the ducky'. Two of the fillers

TABLE 1
Sample stimuli for Experiment 1

Condition	Stimulus sentences
SAME-GENDER	Puppy is having lunch with Panda Bear. He wants some milk.
DIFFERENT/first-mention	Puppy is having lunch with Froggy. He wants some milk.
DIFFERENT/second-mention	Puppy is having lunch with Froggy. She wants some milk.

Notes: Puppy and Panda Bear are male; Froggy is female.

referred to the first-mentioned character, and three to the second-mentioned character. These trials were used as diagnostics of the child's ability to pay attention and perform the task; three or more errors resulted in the exclusion of that child from the analysis. On these trials, if the child responded incorrectly, the experimenters corrected him or her, and played the story again.

The experimental and filler items were pseudo-randomly ordered and combined into 3 lists, with forward and backward versions. There were also two versions of each list, one that placed the first-mentioned character always on the child's right side, and one on the left side. The stories were digitised and played through PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) running on a Power Macintosh.

Monitoring eye movements. In addition to children's responses, we also recorded a close-up video of their face. From this we coded the direction of their gaze during the pronoun and immediately after. A video camera was placed across the table from the child, and was trained on the child's face only. The characters on the table were out of view of the camera, thus enabling the coder to be blind to the experimental condition while coding 'left' or 'right' gaze direction (i.e., toward character 1 or 2). The sound was also turned off during the coding of eye movements so the coder wouldn't know the condition of that trial. The signal from the videocamera was sent to a frame-accurate Sony DSR-30 digital VCR, which recorded 30 images per second. Sound was recorded through a microphone that was connected directly to the digital VCR, yielding frame-accurate sound.

Adult controls. We also investigated the adult response to this task, by having 12 adults participate in an off-line experiment. The stories for the adults followed the same format, and in most cases were identical to, the stimuli for the child experiment. The stories were recorded by the same speaker, using the same prosody, tempo, and rhythm. Unlike the child experiment, characters were presented as pictures on a computer screen. The story played as each visual display appeared, and after the story, the question (Who wants the ball?) appeared at the bottom of the screen. Adults answered the question by clicking on one of the two characters.

Before the experiment, subjects were introduced to the four characters (Puppy, Panda Bear, Froggy, and Bunny) on an introduction screen. They were asked to click on each picture to 'hear something about the character'. These descriptions mentioned the character's name and something about him or her, e.g., 'This is Puppy. He has brown fur'. Each description crucially used a pronoun, which established the characters' gender. In addition, the

pictures depicted characters with clothing consistent with male and female stereotypes.[5]

Results

Adult controls.   As expected, adults generally chose the gender-matched pronoun in the different-gender conditions, and showed a preference for the first-mentioned character in the same-gender condition. In the different-gender/first-mention condition, all participants chose the first-mentioned character 100% of the time. In the different-gender/second-mention condition, they chose the second-mentioned character most of the time. This condition was not error-free, however – participants still chose the first-mentioned character 23% of the time (SE = 6.5%), showing a general first-mention bias. In the same-gender condition, they chose the first-mentioned character 88% of the time (SE = 7.8%), also showing a first mention bias.

These results (Figure 2) are consistent with evidence that adults use both gender and order-of-mention to guide their on-line interpretation of pronouns. The relatively high number of errors in the different/second-mention condition is likely to stem from the fact that the characters were new to the adults, and we did not overly emphasise their genders before the experiment began.

Analyses of variance over arcsine-transformed participant and item means showed a main effect of condition, $F_1(2, 20) = 60$, $p < .001$; $F_2(2, 22) = 59$, $p < .001$. There was no significant effect of the location of the first-mentioned character: right/left: $F_1(1, 10) = 1.5$; $F_2(1, 11) = 1.7$. Crucially, responses in all three conditions differed from chance, where chance is 50%; (different/first-mention: no variation, 100% first-mention responses; different/second-mention: $t(11) = 3.5$, $p < .01$; same-gender: $t(11) = 5.7$, $p < .001$. Thus, adults interpret pronouns systematically, even when gender does not uniquely identify the correct referent, as in the same-gender condition.

Children's off-line results.   Children's initial responses to the question were analysed. In order for a response to count, the child had to completely place the object in front of a character; if they initially reached towards a character and then moved the object to the other character, we only counted where they actually placed the object. Twelve children (five younger and six older) changed their response after the initial response (4% of all the data, $n = 15$). When younger children changed their initial response, the change

---

[5] Twelve items were rotated though the three conditions (same, different-gender/first-mention, and different-gender, second-mention) and pseudo-randomly combined with 12 fillers into 3 lists, with forward and backward versions. Each list began with 2 practice items. Each list had two versions, one with the first-mentioned character always on the right, and the other with the first-mentioned character always on the left.

was equally likely to make their final response more adult-like (57%) as not. Older children's changes were somewhat more likely to make their response more adult-like (75%). If a child answered verbally and didn't move any object, we counted the verbal response.

We only included a child's data in the analysis if he or she performed above chance on the filler trials, where the critical sentence used the characters names, rather than pronouns. On two of the fillers the crucial sentence named the first-mentioned character, and on three it named the second-mentioned character. We excluded two children from the analysis for failing this test. Importantly, the children who were included performed equally well when the name referred to the first- and second-mentioned character ($ts < 1.5$), as shown in Table 2.

In contrast with the adults, children showed no tendency to prefer the first-mentioned character in the same-gender condition. Figure 2 plots the percentage of character 1 responses. That is, the 'correct' response should be close to 100% for the first-mentioned condition, and close to 0% for the second-mentioned condition. This graph shows that children in both age groups were at chance in the same-gender condition. By contrast, children were adult-like in their ability to use gender to guide their interpretation of the pronoun. This was especially strong for the older age group (4;1–5;0), who chose the gender-matched character at a greater than chance rate in both different/first-mentioned and different/second-mentioned conditions. The younger children successfully chose the gender-matched character when it was also the first-mentioned character, but made more errors in the different-gender/second-mention condition. A separate analysis of boys and girls reveals that the younger girls responded like the older children, i.e., responding correctly in the different-gender conditions, and the younger boys responded correctly only in the different/first-mention condition.

We evaluated children's performance by conducting planned t-tests to determine if the average number of first-mention responses in each condition differed from chance, where we calculated responses over both participants and items. The older children responded at a greater than chance level in both different-gender conditions: different/first-mention: $t_1(22) = 10.98$, $p < .001$; $t_2(8) = 13.63$, $p < .001$; different/second-mention: $t_1(22) = -7.37$, $p <$

TABLE 2
Average percentage correct responses on filler items

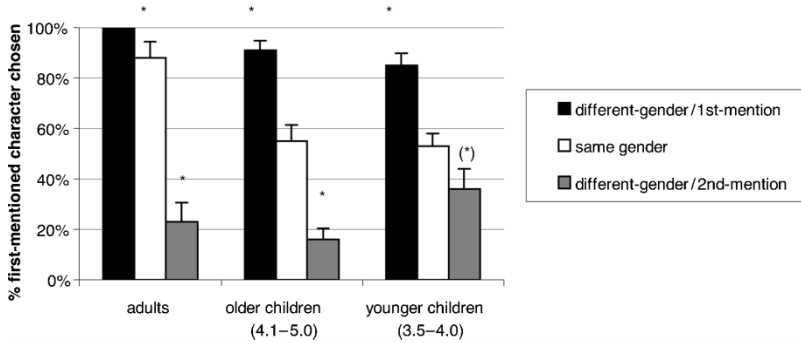|	First-mention fillers	Second-mention fillers
Younger children (3.5–4.0 years)	80%	82%
Older children (4.1–5.0 years)	84%	84%

Figure 2. Response results from adult controls and children in each age group, plotted as the percentage of items on which the participant chose the first-mentioned character. $* = p < .05$; $(*) = 1.0 > p > .05$.

.001; $t_2(8) = -6.89$, $p < .001$. Their behaviour in the same-gender condition, however, was at chance, $t_1(22) = 0.781$; $t_2(8) = 0.89$. The younger children were above chance in only the different-gender/first-mention condition, $t_1(23) = 7.76$, $p < .001$; $t_2(8) = 8.58$, $p < .001$. In the different-gender/second-mention condition they were marginally above chance, $t_1(23) = -1.86$, $p = .08$; $t_2(8) = -2.29$, $p = .05$, and in the same-gender condition they were at chance, $t_1(23) = 0.57$; $t_2(8) = 0.31$. Within the younger group, the girls performed like the older children, responding above chance in both different-gender conditions: different/first-mention: $t_1(11) = 5.61$, $p < .001$; $t_2(8) = 5.97$, $p < .001$; different/second-mention: $t_1(11) = -2.464$, $p < .05$; $t_2(8) = t_2(8) = -3.411$, $p < .01$, and at chance in the same-gender condition, $t_1(11) = -0.364$; $t_2(8) = -0.206$. The boys, by comparison, only responded above chance in the different/first-mention condition, $t_1(11) = 5.138$, $p < .001$; $t_2(8) = 5.66$, $p < .001$, and were at chance in the different/second-mention, $t_1(11) = -0.47$, $t_2(8) = -0.69$, and same-gender conditions, $t_1(11) = 0.14$; $t_2(8) = 1.16$.[6]

---

[6] Analyses of variance supported the above pattern, establishing that children performed differently by condition. Analyses of the data for both older and younger children show a main effect of condition: older: $F_1(2, 42) = 49.06$, $p < .001$; $F_2(2, 16) = 80.58$, $p < .001$; younger: $F_1(2, 44) = 19.40$, $p < .001$; $F_2(2, 16) = 13.49$, $p < .001$, and no effects of the child's sex: older: $F_1(1, 21) = 0.21$; $F_2(1, 8) = 0.03$; younger: $F_1(1, 22) = 1.41$; $F_2(1, 8) = 2.0$, or interactions between sex and condition: older: $F_1(2, 42) = 0.012$; $F_2(2, 16) = 0.03$; younger: $F_1(2, 44) = 0.80$; $F_2(2, 16) = 1.10$. The effect of condition also holds if we only consider the different-gender conditions: older: $F_1(1, 21) = 120.98$, $p < .001$; $F_2(1, 8) = 211.79$, $p < .001$; younger: $F_1(1, 22) = 25.57$, $p < .001$; $F_2(1, 8) = 35.25$, $p < .001$. However, for the younger group only, if we consider only the different-gender conditions, there is an interaction between condition and participants' gender in the items analysis, $F_1(1, 22) = 1.02$, $p = .32$; $F_2(1, 8) = 7.84$, $p < .05$.

Children's eye movement results.   The focus of the current research is to understand not only what children know, but also how they recruit multiple sources of information during on-line language comprehension. We therefore analysed children's eye movements as they heard the pronoun and immediately after, as a measure of the time course over which gender affects their interpretation of the pronoun. The results showed that for the older children, who consistently chose the gender-matched referent, their eye movements paralleled their off-line responses. The younger children, by contrast, had a baseline tendency to fixate the last thing mentioned, i.e., the second character. For this age group, there was little correspondence between their eye movements and their off-line responses.

We examined the video record of the child's face for each 33-ms frame on the digital video tape. We coded whether the participant was looking at the character to the right or the left. We began coding at the onset of the pronoun in the story, and continued for a minimum of 3 seconds, or until the child began to pick up the object that was presented by the experimenter. Other potential fixation categories were (a) experimenter 1, (b) experimenter 2, (c) centre (usually the toy), (d) else, or (e) track loss. Track loss occurred primarily when the child blinked, and accounted for 3% of the data overall. A subset of items was coded by two experimenters to assess inter-coder reliability with this coding system. This comparison showed 87% agreement on the location fixated, and ± 0.4 frames agreement for fixation onset.

The only reliable differences between conditions occurred for the 4–5-year-old children. As Figures 3–6 show, children in this age group were more likely to fixate on the target character in the different-gender conditions, beginning between 600 and 800 ms after the onset of the pronoun. This segment on average corresponded with the onset of the final word in the story. Note that the percentage of fixations on both characters decrease over time; this is due to the structure of the task, which leads children to look at the centre towards the toy that experimenter 1 produces after the story is over.

In the same-gender condition, older children's eye movements also foreshadowed their off-line response. Even though the referent was not uniquely identifiable by gender, children tended to fixate more on the character they ended up choosing. In fact, in this condition the child's initial point of fixation may have helped determine how they interpreted the pronoun. When the first-mentioned character was fixated at the onset of the pronoun, children chose it 78% of the time, but when the second-mentioned character was fixated at pronoun onset, they only chose the first-mentioned character 48% of the time ($\chi^2 = 3.6$, $p = .057$).

By contrast, the younger children (both girls and boys) showed no preference for fixating on the target, or on the character they would eventually choose in the same-gender condition. Instead, younger children
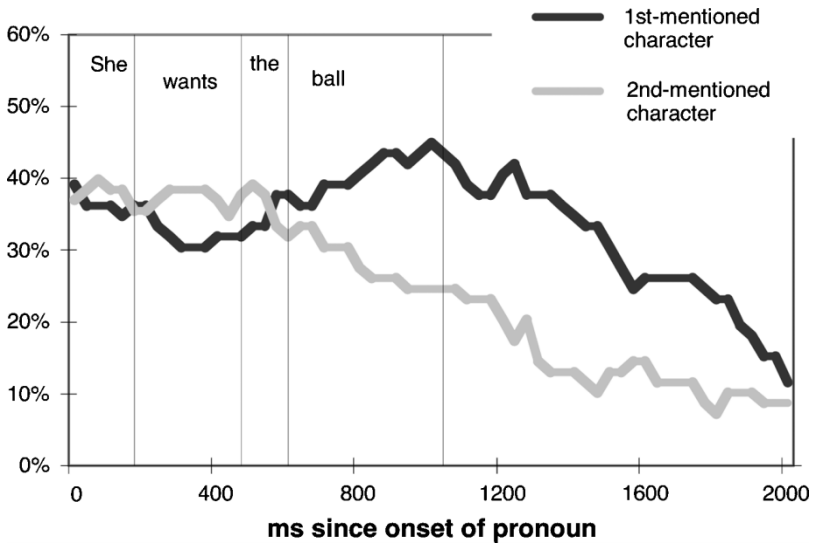
Figure 3. Eye movement results for Experiment 1. Older children: Different-Gender/First-Mention.
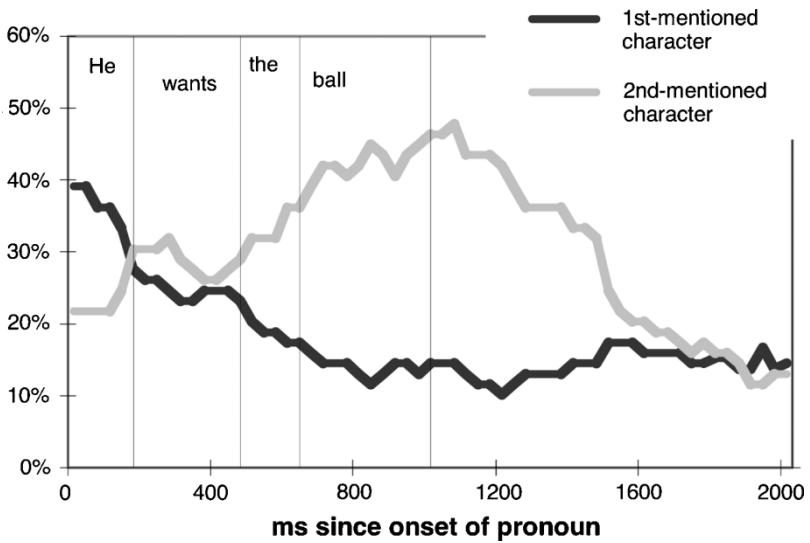

Figure 4. Eye movement results for Experiment 1. Older children: Different-Gender/Second-Mention.
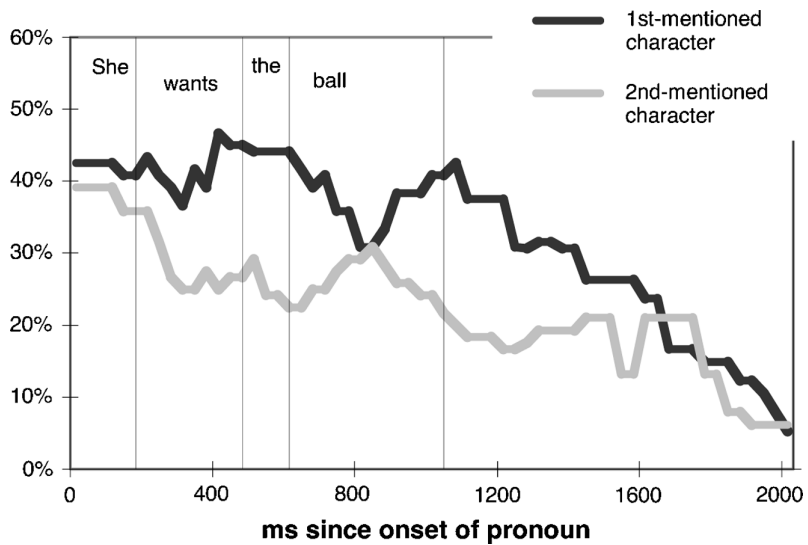
Figure 5. Eye movement results for Experiment 1. Older children: Same gender items where they ultimately chose the first-mentioned character.
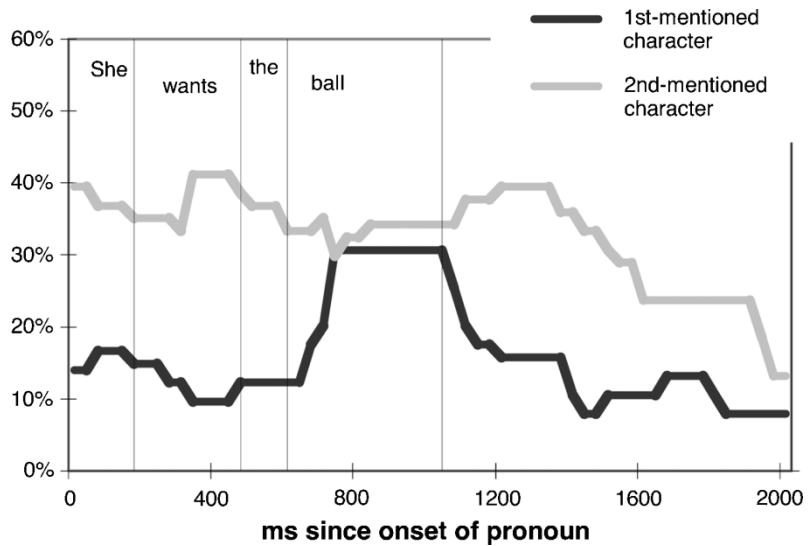

Figure 6. Eye movement results for Experiment 1. Older children: Same gender items where they ultimately chose the second-mentioned character.

had a tendency to be still fixating on the second-mentioned character at the onset of the pronoun, in all conditions. This is likely to be the result of continuing to fixate the most recently mentioned entity, i.e., the second-mentioned character.

We assessed the reliability of eye movement patterns by performing analyses of variance on 'first-mention advantage scores', which were the percent of fixations on the first-mentioned character – percent fixations on the second-mentioned character. The average 'first-mention advantage' for participants and items was calculated for 200-ms segments, beginning at the onset of the pronoun, and continuing until 1600 ms.

The older children's eye movements were influenced by the experimental condition (different/first-mention; different/second-mention; same), starting between 600 and 800 ms after the onset of the pronoun. Analyses of variance revealed no significant effects of condition for the first three segments (Fs < 1, ps > .3). An effect of condition emerged as a marginal effect during the 4th segment (600–800 ms after pronoun onset), $F_1(2, 44) = 2.56$, $p = .089$; $F_2(2, 16) = 3.17$, $p = .069$, and as a robust effect for segments 5–8; segment 5: $F_1(2, 44) = 5.66$, $p < .01$; $F_2(2, 16) = 11.09$, $p < .005$; segment 6: $F_1(2, 44) = 6.23$, $p < .005$; $F_2(2, 16) = 14.20$, $p < .001$; segment 7: $F_1(2, 44) = 6.13$, $p < .05$; $F_2(2, 16) = 8.65$, $p < .005$; segment 8: $F_1(2, 44)$: 4.30, $p < .05$; $F_2(2, 16) = 4.30$, $p < .05$. If we only consider the different-gender conditions, the identical patterns of significance obtain.

For the younger children, there were no significant effects of condition during any of the segments, either considering all data together, or only the different-gender conditions (all Fs < 2, all ps < .2). Instead, the younger children tended to fixate the second-mentioned character at the onset of the pronoun and immediately after: the average first-mention advantage from 0–200 ms after the onset of the pronoun was −.2037 for the younger children, whereas it was .0308 for the older children. This difference emerged in omnibus ANOVAs as an effect of age (3–4-year-olds vs. 4–5-year-olds) that was significant from 0–200 ms after the pronoun onset, $F_1(1, 45) = 7.71$; $p < .01$; $F_2(1, 8) = 8.42$, $p < .05$, and marginal from 200–400 ms after the pronoun, $F_1(1, 45) = 3.42$; $p = .071$; $F_2(1, 8) = 4.16$, $p = .076$.

Discussion

The response results showed that all but the youngest boys were able to correctly interpret pronouns when the gender uniquely identified a referent. This suggests that by age 4, all children are able to use gender to guide pronoun resolution. But none of the children showed the adult pattern of favouring the first-mentioned referent in the same-gender condition. When the two characters had the same gender, children in both age groups were equally as likely to choose the first-mentioned or second-mentioned

character as the referent of the pronoun. The only hint of a first-mention bias emerged in the finding that children, like adults, were more likely to correctly interpret the pronoun when both gender and order-of-mention pointed toward the same referent. This is perhaps evidence that children of this age are beginning to build a sensitivity to the order-of-mention cue, even though they are still unable to use order-of-mention by itself when gender information is not informative.

We additionally saw evidence from 4–5-year-old children that their off-line responses matched the referents they considered as they were listening to the story. This was true for the two different-gender conditions, where children in this age group tended to both look at the gender-matched character during the story, as well as choose that character as the referent of the pronoun.

These results establish two patterns. First, children begin to depend on gender for the interpretation of pronouns earlier than they are able to use order-of-mention. We discuss this finding below in terms of the reliability of gender information, compared with order-of-mention. Second, the eye movement results suggest that by age 4, gender affects children's on-line hypotheses about the referent of the pronoun.

However, this study left open several questions about children's on-line interpretation of pronouns. Children only began to preferentially fixate the target referent between 600 and 800 ms after the onset of the pronoun, which is at least 200 ms later than adults in a similar task (Arnold et al., 2000a). However, this difference may be due to task differences, since the adults in Arnold et al.'s (2000a) task were asked to verify whether the story matched the picture, which may have encouraged closer time-locking of eye movements to the linguistic input. In addition, it is difficult to interpret the eye movements in the same-gender condition of the current experiment, since the pronoun is never disambiguated by the story or visual context, thus making both interpretations equally 'correct'. To address both of these concerns, we examined the on-line processing of pronouns in Experiment 2, using the same task as in Arnold et al. (2000a).

EXPERIMENT 2

The goal of this experiment was to more accurately assess the time course over which children use gender and order-of-mention to guide their consideration of potential referents during pronoun resolution. We used a lightweight head-mounted eyetracking visor (Tanenhaus et al., 1995; Trueswell et al., 1999) to obtain more precise information about fixations than is possible with the video system used in Experiment 1. We used a version of a task developed for adults in an earlier study (Arnold et al.,

2000a). Children viewed a picture like the one in Figure 7, while they listened to a story about the picture, e.g., 'Donald is bringing some mail to Mickey, while a big rain storm is beginning. He's carrying an umbrella, and it looks like they're both going to need it.' The story was narrated by a puppet, and the child's task was to identify whether the puppet made any mistakes – i.e., whether the story matched the picture or not.

This paradigm offered several advantages. It encouraged closer time locking between eye movements and the linguistic input, since the child's gaze followed the description of the picture, as they verified each statement in the story. This paradigm also established the correct referent of the pronoun through the visual context and story, which allowed us to compare the eye movements in both different-gender and same-gender conditions.

We tested children who were slightly older than the children in Experiment 1, since younger children often refuse to wear the eyetracking visor. This further allowed us to examine whether order-of-mention might begin to emerge around age 5.

Method

Participants. We report data from 16 children ages 4;0 to 5;9, average 5;2. Data from an additional nine children had to be replaced because they
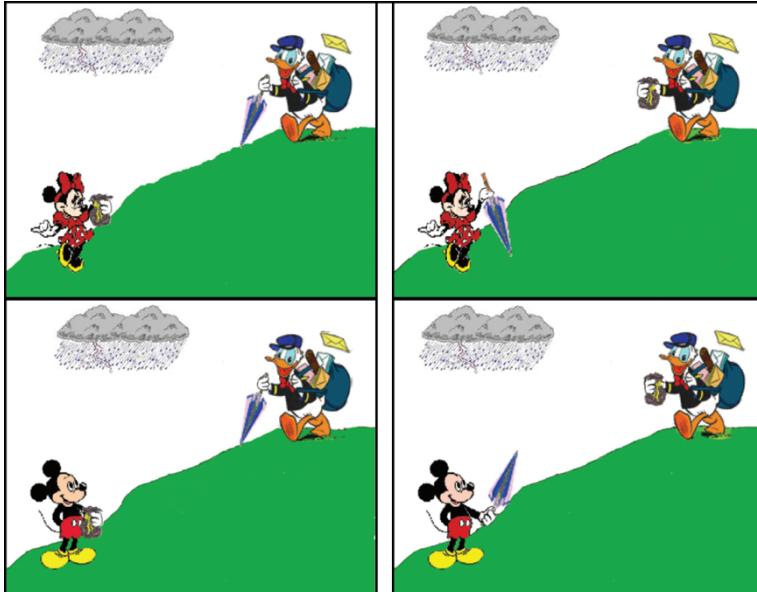

Figure 7. Sample visual stimuli for Experiment 2. Adapted from Arnold et al. (2000a).

failed the diagnostic trials ($n = 4$), were unable to do the task ($n = 1$), were non-native speakers of English ($n = 2$), or stopped in the middle of the experiment ($n = 2$). Six additional children participated without producing any data, since we were unable to calibrate them on the eyetracker. Twenty-one children were recruited and tested in daycare centres in Philadelphia, and 10 children were tested in the laboratory at the Institute for Research in Cognitive Science at the University of Pennsylvania.

Eight children were assigned to the different-gender condition (average age 5;4; 6 boys, 2 girls) and eight were assigned to the same-gender condition (average 5;0, 6 girls, and 2 boys).

Procedure.   Children's eye movements were monitored as they viewed a picture like those in Figure 7, and listened to a story like those in Table 3. Their task was to verify whether the picture matched the story. Eye movements were monitored with an ISCAN head-mounted eyetracker (for eyetracking details, see Trueswell, et al., 1999). The point of fixation was recorded as a crosshair, superimposed over a video record of the scene that children were viewing. Scene and fixation information were recorded once for each 33-msec frame of video tape, using a digital VCR (a SONY DSR-30). The child sat on a child-sized chair and viewed each picture on a laptop computer that was on a small table.

The principal characters in each scene were cartoon characters that are familiar to most children of this age: Mickey Mouse, Minnie Mouse, Donald Duck, and Daisy Duck. These characters are particularly advantageous for a study about gender, since they have stereotypically male and female appearances, for example the females wear frilly dresses and high heeled shoes.

There were four versions of each story and accompanying picture, following a 2 (same vs. different gender) $\times$ 2 (first vs. second mention) experimental design. Gender was manipulated between subjects, and order-

TABLE 3
Sample auditory stimuli for Experiment 2

Condition	Stimulus sentences
Same-Gender	Donald is bringing some mail to Mickey, while a big rain storm is beginning. *He's* carrying an umbrella, and it looks like they're both going to need it.
Different-Gender	Donald is bringing some mail to Minnie, while a big rain storm is beginning. *He's/She's* carrying an umbrella, and it looks like they're both going to need it.

of-mention within subjects. One version of each item was assigned to each of 4 lists, each having forwards and reverse order.

Each text had four clauses, broken into two sentences, as in Table 3. The first clause mentioned the two characters; the second mentioned some other object in the picture. This clause played the important function of drawing eye movements away from the two characters, with the intention of avoiding any baseline tendency to be fixating the second-mentioned character at the onset of the pronoun. The third clause began with a pronoun referring to one character or the other, and the final clause provided concluding information without mentioning either character individually.

In all cases the picture and story together disambiguated the pronoun, but not until after the verb. For example, in the pictures in Figure 7 both characters are holding something, but only one is holding an umbrella. In half the items the first-mentioned character was on the right of the picture, in half it was on the left.

Texts were digitally recorded to a computer by a female speaker. The same recording was used for all conditions of a given item, with the differing names and pronouns spliced in using SoundEdit 16. The pictures and texts were controlled by PsyScope 1.0.2 running on a G3 PowerBook.

Each experimental session began with the introduction of the two experimenters, the narrator puppet, and the equipment. We demonstrated to the child how the eyetracking equipment showed where the wearer was looking, and asked if the child would like to wear it. If the child consented, we calibrated the tracker with a point-of-light calibration system.

We then explained the task to the child, in which the puppet would tell stories about each picture. At the beginning of all experimental sessions, we verified that participants were familiar with the names of the principal cartoon characters. If they weren't, we practiced the names until we were confident they had learned them.

The picture-verification task took the form of asking the child to correct the narrator puppet, who was presented as likely to make mistakes. This was illustrated by asking the puppet two simple questions, one of which he answered incorrectly. When the puppet said correct things, the child gave him a piece of candy, and when he said incorrect things, the puppet was required to do push-ups (or some similar punishment suggested by the child). This game engaged children in the task enthusiastically.

We practiced the task with the child with an initial warm-up story, which was performed by one of the experimenters, and included a blatantly incorrect statement. After that, the puppet hid under the table during each story, and his 'voice' came out of the speakers that were placed on either side of the laptop. The first two recorded stories were practice items, followed by 8 critical items (all 'correct'), combined with 8 fillers (3 'correct', 5 'incorrect').

Results and discussion

Overall there was good evidence that children understood the stories. Children tended to correctly accept the puppet's story on the experimental items (mean = 86%). Occasionally they identified an unintended 'mistake', e.g., when the story said 'Mickey is reading a story to Donald/Daisy, under a big tree outside', three children said the tree wasn't big. However, the rate of 'no' answers did not differ by condition (all Fs < 3, all ps > 1). Since it is often more difficult for children to appropriately reject a story, we used their 'no' answers on filler items as a diagnostic of their ability to pay attention and do the task. Children were only included in the analysis if they correctly answered 2/5 of the 'no' fillers (i.e., if they said that the puppet had made a mistake on at least 2 'incorrect' trials). Those children that were included in the analysis correctly said 'no' on an average of 73% of the 'no' fillers.

We examined children's eye movements for the period of time during and immediately after the pronoun. We coded these fixations by examining the video record of the scene, with a superimposed fixation point, for each 33-ms frame on the videotape. Point of fixation was coded in four categories: target (the referent of the pronoun), competitor (the other character), other (something else), or track loss. Track loss occurred mostly when the child blinked, and accounted for 8% of the data.

Figures 8a and 8b show the average proportion of fixations on the target, competitor, and other at each frame. These results again reveal a robust use of gender for pronoun comprehension. In the two different-gender conditions, children began converging on the correct referent beginning 400 msec after the onset of the pronoun. This rapid speed of pronoun identification is the same as that observed for adults doing the same task (Arnold et al., 2000a; see Figure 9).

By contrast, these children did not use order-of-mention to consistently guide their interpretation of the pronoun. In the same-gender conditions, when gender did not uniquely identify a referent, children did not converge on the target until well after the point of disambiguation, e.g., the word 'umbrella' in the example in Table 3. There were numerically more early fixations on the target when it was the first-mentioned referent, suggesting that order-of-mention might be partially guiding their interpretation. This resulted in a trend towards an interaction, such that there were fewer target looks in the same/second-mention condition than in the other three conditions. However, this interaction was not statistically reliable, except for a marginal effect in the participants' analysis between 600 and 1000 ms after the onset of the pronoun. This contrasts with the strong and early effect of gender on children's pronoun comprehension. It also contrasts with the performance of adults doing the same task, who produced a reliable
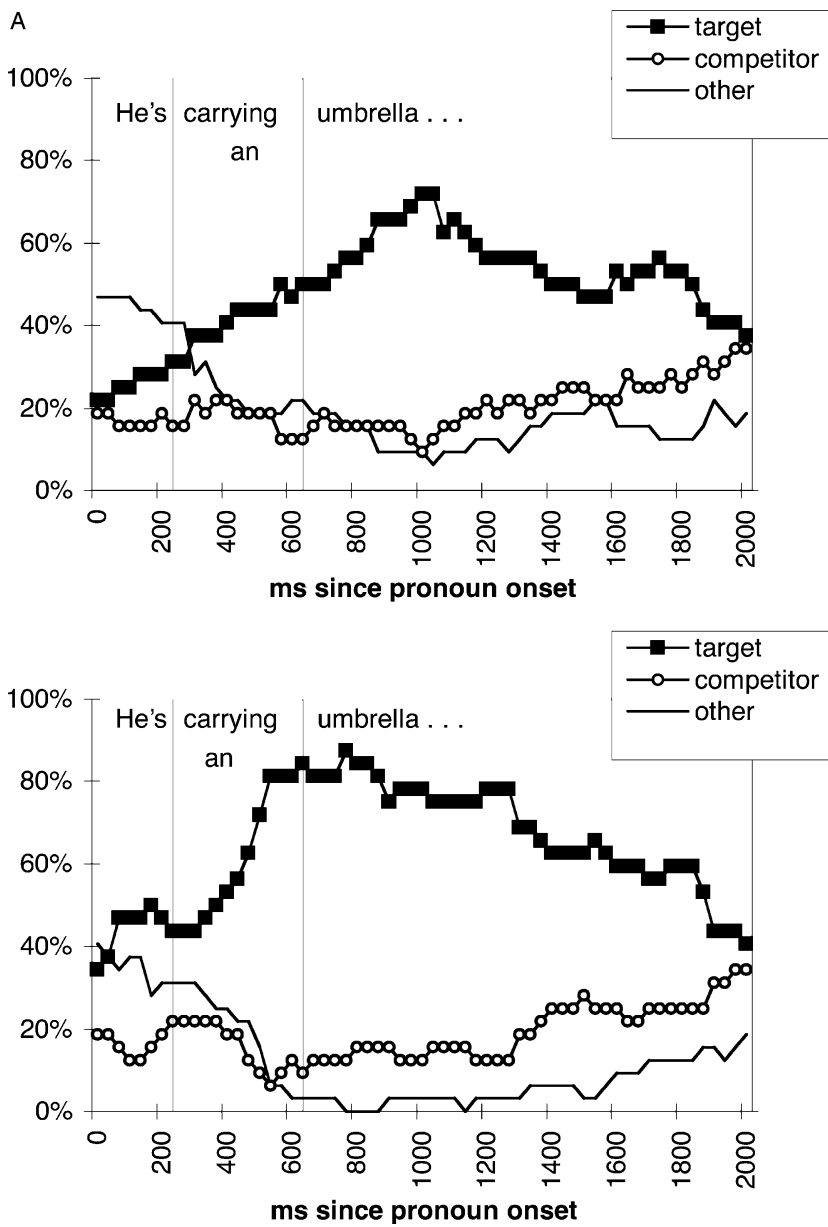
Figure 8a. Eye movement results for Experiment 2. Different-Gender/First-Mention condition (top panel); Different-Gender/Second-Mention condition (bottom panel).
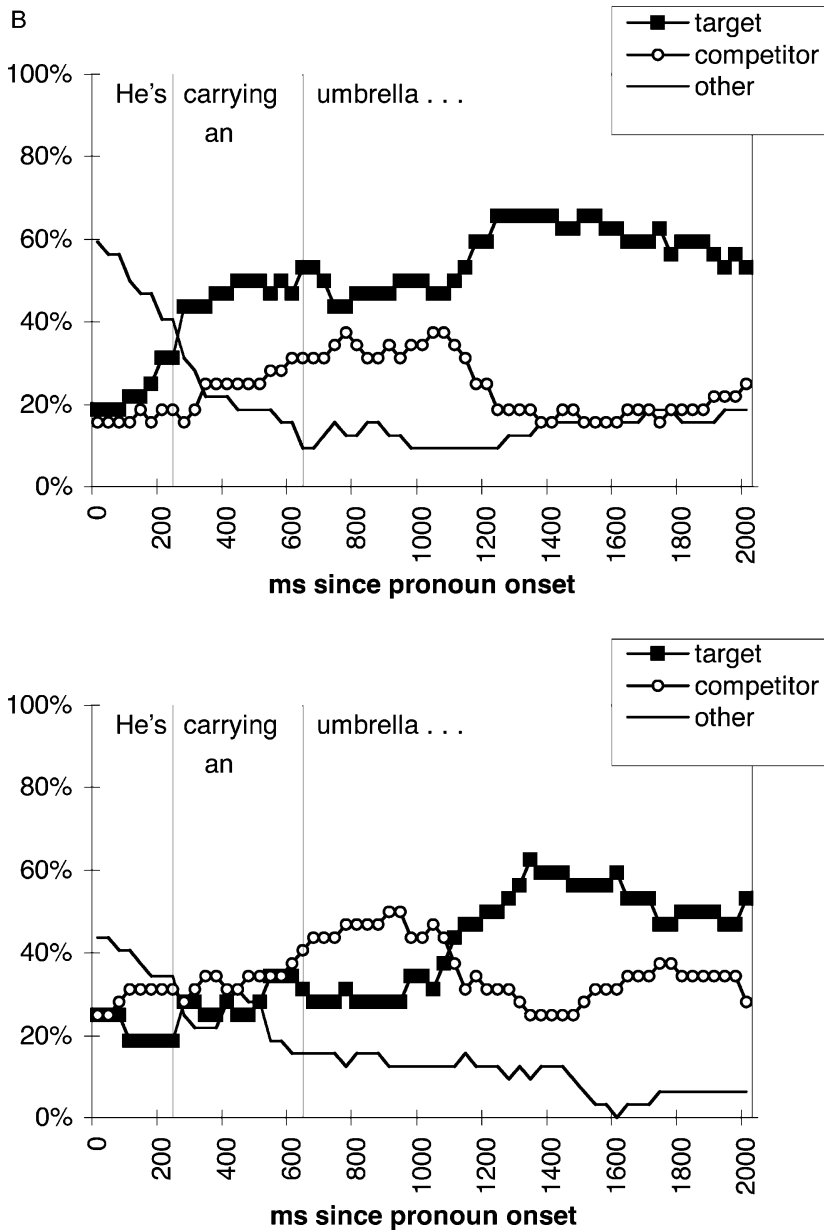
Figure 8b. Eye movement results for Experiment 2. Same-Gender/First-Mention condition (top panel); Same-Gender/Second-Mention condition (bottom panel).
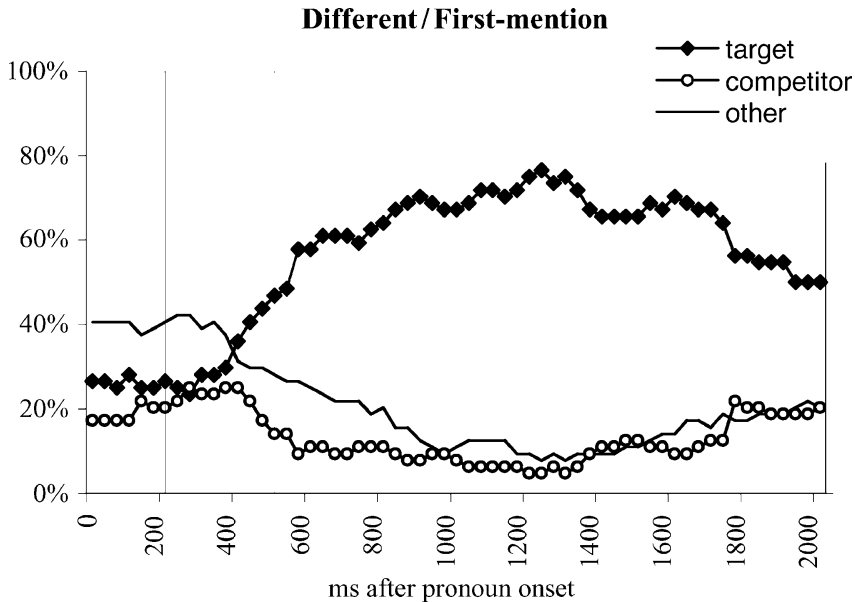
553

Figure 9.   Adults eye movements in the Different/First-mention condition, shown as percentage of looks to each object, starting at the onset of the pronoun. Data also published as Figure 2 in Arnold et al. (2000a).

interaction between gender and referent much earlier, before the point of disambiguation (Arnold et al., 2000a).

We tested the reliability of these findings by examining the 'target advantage', i.e., the proportion of fixations on the target minus the proportion of fixations on the competitor. Analyses of variance were performed over participant and item mean target advantages for eight 200-ms windows, beginning at the onset of the pronoun. The first two segments showed no main effects or interactions. From 400–1200, there was a main effect of gender, corresponding to the fact that subjects looked more at the target than competitor in the different-gender conditions, but not in the same-gender conditions. The same segments showed no effect of referent. Gender and referent also did not interact, except for a marginal interaction in the participants' analysis only, from 600–1000 ms after the pronoun (see Table 4 for statistics).

In sum, the eyetracking results revealed a strong use of gender for the interpretation of pronouns, which occurred as early as for adults in the same task. By contrast, there was no reliable evidence that 5-year-old children preferred first-mentioned characters as the referent for ambiguous pronouns.

Nevertheless, our data provide a hint that children at this age may be on the verge of acquiring sensitivity to order-of-mention, or may be very weakly

TABLE 4
The reliability of the Gender Effect, Referent Effect, and Interaction (Gender $\times$ Referent)

Segment	Effect of Gender	Effect of Referent	Interaction ($G \times R$)
0 to 200 ms	n.s.	n.s.	n.s.
	($F_1 = 1.46$; $F_2 = 1.37$)	($F_1 = 0.65$; $F_2 = 0.01$)	($F_1 = 1.77$; $F_2 = 1.45$)
200 to 400 ms	n.s.	n.s.	n.s.
	($F_1 = 0.96$; $F_2 = 0.42$)	($F_1 = 0.39$; $F_2 = 0.81$)	($F_1 = 1.36$; $F_2 = 1.16$)
400 to 600 ms	Reliable, ps $< .05$	n.s.	n.s.
	($F_1 = 10.20$; $F_2 = 5.86$)	($F_1 = 0.32$; $F_2 = 0.08$)	($F_1 = 2.17$; $F_2 = 1.55$)
600 to 800 ms	Reliable, ps $< .001$	n.s.	F_1 marginal (3.64)
	($F_1 = 21.60$; $F_2 = 38.36$)	($F_1 = 0.12$; $F_2 = 0.01$)	F_2 n.s. (2.46)
800 to 1000 ms	Reliable, ps $< .001$	n.s.	F_1 marginal (3.75)
	($F_1 = 21.43$; $F_2 = 26.98$)	($F_1 = 0.73$; $F_2 = 0.97$)	F_2 n.s. (0.07)
1000 to 1200 ms	Reliable, ps $< .005$	n.s.	n.s.
	($F_1 = 12.49$, $F_2 = 27.68$)	($F_1 = 0.26$; $F_2 = 0.81$)	($F_1 = 1.33$; $F_2 = 1.35$)
1200 to 1400 ms	n.s.	n.s.	n.s.
	($F_1 = 1.22$; $F_2 = 0.81$)	($F_1 = 0.83$; $F_2 = 0.27$)	($F_1 = 1.96$; $F_2 = 2.40$)
1400 to 1600 ms	n.s.	n.s.	n.s.
	($F_1 = 0.03$; $F_2 = 0.25$)	($F_1 = 0.33$; $F_2 = 0.06$)	($F_1 = 1.01$; $F_2 = 1.68$)

Notes: Analyses are performed at each 200-ms segment of time, with respect to the proportion looks to the target minus the proportion looks to the competitor. All F_1 ANOVAs have df (1, 14), and all F_2 ANOVAs have df (1, 7).

sensitive to it. We already observed that the eye movement data show a trend (though statistically unreliable) towards more target looks in the same/first-mentioned condition, compared with the same/second-mention condition. In addition, children's off-line responses revealed a first-mention bias for three of the girls in the same-gender condition. Recall that the child's task was to say whether the narrator puppet had made any errors, and if so to identify what the error was. Occasionally these explanations of the error revealed that the child had taken the pronoun to refer to the first-mentioned condition, when the picture required a second-mentioned interpretation. For example, when the story was 'Mickey is reading a story to Donald . . . He's sitting on a rock', the second-mention condition pictured Donald on a rock and Mickey on a log. One child said 'but Mickey isn't on a rock'. This type of response was produced by three girls, but only on one or two of the four items in the same-gender/second-mention condition.

It appears that these three girls may have begun to follow the adult-like strategy of assigning pronouns to the first-mentioned character of the preceding sentence. However, their answers suggest that they may have applied this strategy ballistically, and were unable to accept a second-mentioned interpretation of the pronoun. This contrasts with adults, who

accept pronouns for second-mentioned characters, albeit less readily than for first-mentioned characters (e.g., Arnold et al., 2000a).

Despite the off-line appearance of a first-mention strategy in these three girls, their eye movements do not show a clear use of order-of-mention for on-line processing. If they had, we might have expected them to converge on the target soon after the pronoun in the same/first-mention condition, but not the same/second-mention condition, as adults did. Figure 10 shows that the small first-mentioned preference emerged relatively late, around 500–600 ms after the onset of the pronoun. Although this falls before the *average* onset of the disambiguating word, it comes after disambiguating information about the pronoun referent was available in some items.

In addition, it is important to bear in mind that most children showed little evidence of any bias toward the first-mentioned character. In fact, two children (both boys) also provided a 'second-mention bias' response in one item in the same/first-mention condition,[7] for example saying that the above-mentioned story was wrong because Donald was on a log.

GENERAL DISCUSSION

In two experiments, we found that young children are quite good at using the gender information on English pronouns to identify a referent. By age 4, children correctly interpret pronouns when there is only one gender-matched referent in the discourse, and eyetracking evidence from 5-year-olds reveals that they use gender information as rapidly as adults do. By contrast, there was limited evidence of the adult tendency to assign pronouns to referents that had been mentioned first in the previous sentence. In Experiment 1, both 3-year-olds and 4-year-olds were at chance when more than one character matched the gender of the pronoun. The only suggestion of any first-mention bias emerged in the different-gender conditions, where younger children showed fewer errors in the first-mention than second-mention conditions. In Experiment 2, 5-year-olds showed no reliable tendency to look at the first-mentioned character when the pronoun was ambiguous, until well after the pronoun became disambiguated by the story and visual context. The only hint of a first-mention bias in the eye movement data emerged later than the gender effect, and was not statistically reliable. We also observed evidence of a first-mention bias in the off-line responses of 3 subjects in

---

[7] One of these boys also identified an error in one of the same/second-mention conditions that may have been a first-mention bias response. This answer could not unequivocally be characterised as such, and this, together with the fact that he also produced a second-mention response, excluded him from the 'first-mention bias' group.
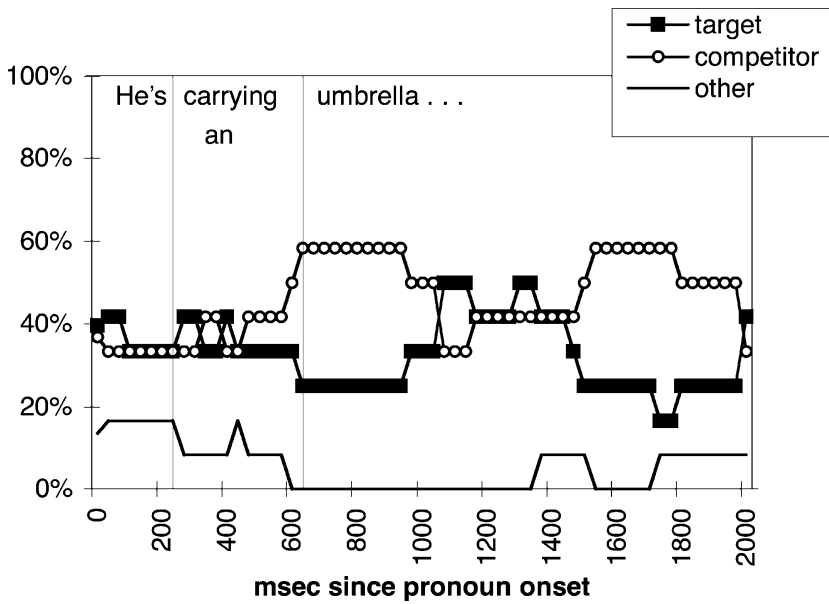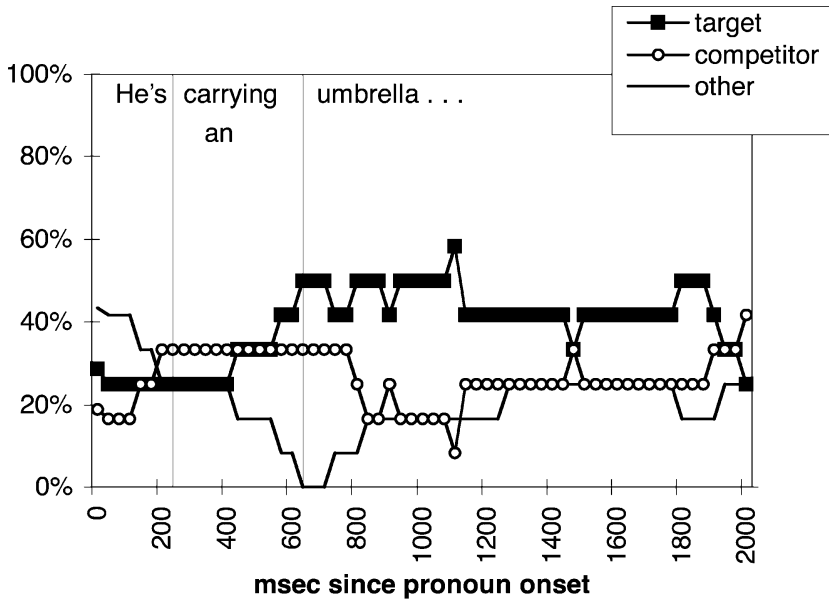
Figure 10. Eye movement results from the three girls who showed off-line evidence of using a first-mention strategy. Same-Gender/First-Mention condition (top panel); Same-Gender/Second-Mention condition (bottom panel).

Experiment 2, but this effect didn't guide their on-line consideration of potential referents for the pronoun.

Our findings seem to contrast with the results of other on-line studies of pronoun comprehension in children, with respect to the use of both gender and accessibility-based information. On one hand, we found that 4 and 5-year-olds used gender to guide both on- and off-line pronoun processing, whereas Tyler (1983) concluded from her on-line experiments that 5-year-olds are poor at using lexical constraints for pronoun resolution. However, as we mentioned above, she did find some ability for 5-year-olds to use gender information for pronoun comprehension, compared with number information. Furthermore, the discrepancy between our findings and Tyler's may be due to the experimental tasks used. Tyler's mispronunciation task, while useful for tapping on-line processes, may have focused children's attention on the surface form of the stories, rather than their meaning, which may have made it relatively difficult for the children to access meaning-related features of characters, like gender. Children in the present experiments were provided with a visual context for the stories, with stereotypically gendered features. This may have freed up memory resources to focus on the meaning of the stories, in addition to supporting a rich representation of the characters' genders. Children typically use language for co-present objects, in highly concrete and contextualised situations, so mimicking these character-istics in the experimental setting offers the greatest chance of observing a child's linguistic abilities.

Our findings also contrast with Song and Fisher's (2005) conclusions that 3-year-olds focus on first-mentioned characters, and use this for on-line pronoun resolution. They claim that children naturally focus on the referents of subject nouns, due to the primacy of the semantic roles that tend to inhabit subject position, and a tendency to remember first-mentioned information better. However, our results show that even older children do not show a subject/first-mention bias, either on-line or off-line. This is likely to be the result of differences in our stimuli. We tested children's ability to use order-of-mention in a single clause, whereas Song and Fisher's stories contained two or more clauses, in which the same character was mentioned first. Even in these circumstances, they observed effects of accessibility much later than the gender effects in our second experiment, which guided children's fixations as early as 400 ms after the onset of the pronoun. Thus, Song and Fisher's results suggest that young children do know that pronouns should refer to accessible entities, even if they do not employ that information as rapidly as adults do. However, our results suggest that order-of-mention alone does not constrain accessibility.

Instead, our findings are generally consistent with other research about language processing in young children, which suggests that children initially recruit information from highly reliable sources, and only later come to

depend on probabilistic and less available information (Trueswell & Gleitman, 2004). Gender provides an extremely reliable cue to the pronoun's referent – 'he' is nearly always used for males, and 'she' for females. Evidence suggests that children can distinguish males from females by age 2.5 (Martin, 1993; Fagot & Leinbach, 1993). Therefore, gender information is available as soon as children learn to map the linguistic feature onto that semantic feature. This is a general process they already have ample experience with, in the domain of mapping lexical items to their referents.[8]

In comparison, children did not show strong evidence of using order-of-mention, either on-line or off-line. These results make sense if we remember that order-of-mention is not simply a cue to pronoun resolution, but rather is a potential tool for establishing joint attention with the speaker. Although the speaker may have mentioned two characters, the order of their mention indicates which one the speaker deems more central to the discourse. Thus, order-of-mention tells listeners who to pay attention to, and it is the resulting accessibility that facilitates pronoun interpretation.

This process poses numerous challenges to the young learner. Even though young children are sensitive to joint attention for the purpose of word learning (e.g., Baldwin, 1991, 1993), the use of finer-grained information to compute distinctions in discourse accessibility may take longer to develop. In the meantime, they may attempt to link pronouns with entities that are in their own capricious focus of attention, whether or not such a link is supported by the discourse context. If this happens, there will be many cases in which they will successfully interpret the pronoun – i.e., those cases where a single entity is under discussion, and the child's attention is focused on that entity as a result of participating in the discussion. This would explain how young children succeed in cases where multiple sources of information support a single character as the more accessible one, as in Song and Fisher's (2005) experiments, but do not when there is more than one relatively accessible entity, as in the results reported here. Our first experiment further supported this idea, where younger children succeeded most often when the gender-matched character was also first mentioned. It would also be consistent with the tendency for the older children in Experiment 1 to interpret gender-ambiguous pronouns as coreferential with the entity they were fixating on at the time of the pronoun. But an egocentric strategy will also lead to many cases of failed pronoun resolution. This may lead children to view accessibility as a weak and unreliable

---

[8] It is notable that the use of gender in language comprehension may in some cases precede its use in production. Some children in Experiment 1 made production errors during the experiment (e.g., 'He a girl' in reference to Froggy), or were reported to make errors by their parents. Nevertheless, most children in this experiment showed evidence of correctly interpreting gender-disambiguated pronouns off-line.

constraint on pronoun identity, and consequently assign little weight to this source of information (Trueswell & Gleitman, 2004).

At some point a child will realise that it is important to pay attention to the public discourse record, which is a reliable indicator of what will continue to show up in the discourse.[9] Among those things that have been mentioned, first-mentioned things tend to be those that the speaker will mention again. Indeed, production evidence shows that adult speakers place information that is more accessible to them early in the sentence (Arnold et al., 2000b; Bock, 1982, 1986; Bock & Irwin, 1980; Ferreira & Dell, 2000; Ferreira, 2003). The expectancy hypothesis suggests that children observe how speakers treat first-mentioned information in the subsequent discourse (Arnold, 1998). They can infer that something is salient if the speaker refers back to it in the next utterance (with any kind of expression, not just a pronoun), if they evoke that information indirectly, or even if they make reference nonlinguistically, e.g., with a gesture. Over time, a child will learn that first-mentioned characters are continued in a discourse more often than second-mentioned characters, and thus associate salience with this position. Yet this is a probabilistic process that will be difficult for a young learner, who has limited exposure to sentences with more than one referential entity. Furthermore, the accessibility of first-mentioned entities modulates with respect to other factors, for example the lexical semantics of the verb (e.g., Arnold, 2001; Stevenson et al., 1994).

This may seem like an insurmountable challenge to achieving an adult-like processing system. Fortunately, the child is not facing the problem of pronoun comprehension in isolation. The need to coordinate a joint discourse model is relevant to almost every aspect of successful communication. Even referring expressions with more lexical content, like 'the dog', need to be defined with respect to a particular domain of interpretation. Similarly, evidence about referent accessibility does not only come from instances where a pronoun is used. According to the expectancy hypothesis, one of the things that children need to learn is that first-mentioned things are more likely to be central to the current discourse segment than those produced later in an utterance. One clue about this association comes from observing what the speaker continues to talk about, whether or not the speaker uses a pronoun. Consider (2), an excerpt from a children's book, 'The Wizard of Oz' (Baum, 1900/1965):

---

[9] Learning the correlation between features like order-of-mention and the likelihood of subsequent mention will have the effect of aligning the child's discourse model with that of other conversational participants. It is important to note that this alignment would be approximated even if the child does not explicitly model the minds of their interlocutors.

... he [Toto] was a little black dog, with long silky hair and small black eyes that twinkled merrily on either side of his funny, wee nose. Toto played all day long,

The first sentence mentions several entities: Toto, long silky hair, small black eyes, and his funny, wee nose.[10] The next sentence refers back to Toto, who was mentioned first in the preceding sentence. Even though this reference is done with a name, it supports the generalisation that first-mentioned entities tend to be continued in the discourse. In fact, an analysis of children's stories shows that entities in subject position are mentioned again in the following sentence more often than other entities (Arnold, 1998). In a face-to-face situation, the same information may be conveyed nonlinguistically. Consider the example in (3), from the CHILDES database (MacWhinney, 2000; conversation between Amy (female aged 1;8) and Mother in data from Bates):

*MOT:here # the dog's gon (t)a go up the ladder.
%act: < 1w > picks up dog and makes him climb the ladder.

In this example, the mother mentions the dog in first-mentioned position, and then manipulates the dog physically, which shows the child that she is attending to it.

Therefore, the complex process of coordinating one's focus of attention with other discourse participants is supported by multiple components of discourse structure. While children are still learning these cues, they can depend on alternative cues to salience in a joint discourse model. When one character has been more recently mentioned, or has been mentioned several times, it provides evidence that this character is salient to the discourse. As Song and Fisher (2005) showed, when multiple cues point to a single referent, 3-year-olds can understand gender-ambiguous pronouns. The correlation of the rich information available in natural discourse not only helps children assign pronouns to accessible referents, but may also help them learn that first-mentioned characters are more salient in the discourse.

Meanwhile, as children gather evidence about order-of-mention, they can rely on their knowledge that the word 'he' differs from 'she'. Lexical information like gender can help children learn which entities the speaker is talking about, and thus attending to. The reliability of gender information means that it may be easier for children to recruit rapidly, in order to use it 'on-line', as the utterance is perceived. This leads children to eventually

---

[10] Even though Toto's hair, eyes, and nose are a part of him, they can be treated as separate entities for the purposes of reference. For example, if the author had wanted to focus on the hair, he might have rephrased the sentence to put the hair in first-mentioned position: 'Long silky hair covered Toto's body. It was always getting matted.'

become adults, who show a strong and stable preference to focus on first-mentioned entities.

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. E. (1998). *Reference form and discourse patterns*. (Doctoral dissertation, Stanford University, 1998). *Dissertation abstracts international, 59/08*, 2950.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes, 31*, 137–162.

Arnold, J. E. (2006). *THE BACON* not *the bacon:* How children and adults understand accented and deaccented noun phrases. *Manuscript submitted for publication.*

Arnold, J. E., Brown-Schmidt, S., Trueswell, J., & Fagnano, M. (2005). Children's use of gender and order-of-mention during pronoun comprehension. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 261–281). Boston, MA: MIT Press.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000a). The rapid use of gender information: Evidence of the time course of pronoun resolution from eye tracking. *Cognition, 76*, B13–B26.

Arnold, J. E., & Griffin, Z. (2006). The effect of additional characters on choice of referring expression: Everyone competes. *Manuscript submitted for publication.*

Arnold, J. E., Novick, J. M., Brown-Schmidt, S., Eisenband, J. G., & Trueswell, J. C. (2001). Knowing the difference between girls and boys: The use of gender during on-line pronoun comprehension in young children. In A. H. J. Do, L. Domínguez, & A. Johansen Somerville (Eds.), *BUCLD 25: Proceedings from the 25th Boston University Child Language Development Conference* (pp. 59–69). Boston, MA: Cascadilla Press.

Arnold, J. E., & Tanenhaus, M. K. (in press). Disfluency effects in comprehension: How new information can become accessible. In E. Gibson & N. Perlmutter (Eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000b). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76*, 28–55.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*, 875–890.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723–767.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3–96). Cambridge: Cambridge University Press.

Baum, F. L. (1965). *The wonderful wizard of Oz*. (Rev. ed.). New York: Airmont Publishing Co.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*, 1–47.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bock, J. K., & Irwin, D. F. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*, 467–484.

Boland, J. E., Acker, M. T., & Wagner, L. (1998). *The use of gender features in the resolution of pronominal anaphora* (Cognitive Science Technical Reports No. 17). The Ohio State University Center for Cognitive Science.

Brener, R. (1983). Learning the deictic meaning of third person pronouns. *Journal of Psycholinguistic Research*, *12*, 235–262.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*, 137–167.

Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A centering approach to pronouns. In *Proceedings from the 25th Annual meeting of the Association for Computational Linguistics* (pp. 155–162). Stanford, CA: Association for Computation Linguistics.

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.

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago: Chicago University Press.

Clark, H. H. (1983). Making sense of nonce sense. In G. B. Flores d'Arcais & R. J. Jarvella (Eds.), *The process of language understanding* (pp. 297–331). London: John Wiley & Sons.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language*, *47*, 292–314.

Fagot, B. I., & Leinbach, M. D. (1993). Gender-role development in young children: From discrimination to labeling. *Developmental Review*, *13*, 205–224.

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying 'that' is not saying 'that' at all. *Journal of Memory and Language*, *48*, 379–398.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Philadelphia, PA: Psychology Press.

Garnham, A., Oakhill, J., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes*, *7*, 231–255.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, *5*, 459–464.

Gennari, S., & MacDonald, M. C. (in press). Acquisition of negation and quantification: Insights from adult production and comprehension. *Language Acquisition*.

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

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1–42). Amsterdam: John Benjamins Publishing.

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

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory. and Cognition*, *182*, 266–283.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local discourse. *Computational Linguistics*, *21*, 203–225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions. *Language*, *69*, 274–307.

Hickmann, M., & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: A comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language*, *26*(2), 419–452.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L., & Trueswell, J. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*, 597–626.

Kaiser, E., & Trueswell, J. C. (in press). Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond salience. In E. Gibson & N. Perlmutter (Eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1985). Language and cognitive processes form a developmental perspective. *Language and Cognitive Processes*, *1*(1), 61–85.

MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, *29*, 469–492.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Martin, C. L. (1993). New directions for investigating children's gender knowledge. *Developmental Review*, *13*, 184–204.

McDonald, J., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*, 543–566.

Race, D. S., & MacDonald, M. C. (2003). The use of 'that' in the production and comprehension of object relative clauses. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 946–951). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language*. Chichester, UK: John Wiley & Sons.

Snedeker, J., Thorpe, K., & Trueswell, J. (2001). On choosing the parse with the scene: The role of visual context and verb bias in ambiguity resolution. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 964–969). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*, 473–592.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*, 73–108.

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*(5217), 1632–1634.

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

Trueswell, J. C., & Gleitman, L. (2004). Children's eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 319–346). New York: Psychology Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

Tyler, L. K. (1983). The development of discourse mapping processes: The on-line interpretation of anaphoric expressions. *Cognition*, *13*, 309–341.

Wykes, T. (1981). Inference and children's comprehension of pronouns. *Journal of Experimental Child Psychology*, *32*, 264–278.

APPENDIX A

Experimental stimuli from Experiment 1

A. Primary stimulus stories

1. Puppy is having lunch with Froggy/Panda Bear. He/she wants some milk.
2. Panda Bear is at school with Bunny/Puppy. He/she wants the book.
3. Bunny is playing outside with Puppy/Froggy. She/he wants the ball.
4. Froggy is getting ready for school with Panda Bear/Bunny. She/he wants the toothbrush.
5. Puppy is eating dinner with Bunny/Panda Bear. He/she wants the egg.
6. Panda Bear is making dinner with Froggy/Puppy. He/she wants the cup.
7. Bunny is getting dressed with Panda Bear/Froggy. She/he wants the hat.
8. Froggy is making art with Puppy. She/he wants the glue.
9. Panda Bear is taking a bath with Bunny/Puppy. He/she wants the soap.

B. Diagnostic filler items. The first two items occurred at the beginning of each list and constituted practice items as well as diagnostics of the child's ability to do the task.

1. Panda Bear and Bunny are putting on their socks. Panda Bear wants the blue socks.
2. Froggy is playing doctor with Puppy. Puppy wants a bandaid.
3. Bunny is playing with Panda Bear. Bunny wants the bear sticker.
4. Panda Bear and Puppy are playing in the bathtub. Puppy wants the ducky.
5. Bunny is playing dressup with Froggy. Froggy wants a ring.