ELSEVIER

Cognition 76 (2000) B13–B26

COGNITION

www.elsevier.com/locate/cognit

Brief article

The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking

Jennifer E. Arnold*, Janet G. Eisenband, Sarah Brown-Schmidt, John C. Trueswell

University of Pennsylvania, Philadelphia, PA, USA

Received 10 November 1999; received in revised form 22 January 2000; accepted 1 March 2000

Abstract

Eye movements of listeners were monitored to investigate how gender information and accessibility influence the initial processes of pronoun interpretation. Previous studies on this issue have produced mixed results, and several studies have concluded that gender cues are not automatically used during the early processes of pronoun interpretation (e.g. Garnham, A., Oakhill, J. & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes*, *73* (4), 231–255; Greene, S. B., McKoon, G. & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *182*, 266–283). In the two experiments presented here, participants viewed a picture with two familiar cartoon characters of either same or different gender. They listened to a text describing the picture, in which a pronoun referred to either the first, more accessible, character, or the second. (For example, Donald is bringing some mail to {Mickey/Minnie} while a violent storm is beginning. He's carrying an umbrella….) The results of both experiments show rapid use of both gender and accessibility at approximately 200 ms after the pronoun offset. © 2000 Published by Elsevier Science B.V. All rights reserved.

Keywords: Language comprehension; Pronoun resolution; Gender cues; Accessibility; Referent processing

* Corresponding author. Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, Suite 400A, Philadelphia, PA 19104, USA. Fax: +1-215-573-9247.

E-mail address: jarnold@linc.cis.upenn.edu (J.E. Arnold).

PII: S0010-0277(00)00073-1

1. Introduction

1.1. Establishing the referent of a pronoun

When a speaker uses a pronoun like 'she' or 'he', how does the listener know who the referent is? Many studies have shown that pronoun comprehension is influenced by the accessibility of potential referents (see Arnold, 1998, for a review). In (1), for example, Clinton is more accessible than Gore, by virtue of coming first in the sentence (Gernsbacher, 1989), so comprehenders might have difficulty assigning the pronoun to the correct referent, Gore. In addition to accessibility, it would seem obvious that pronoun comprehension should be guided by the gender information carried on the pronoun. For example, the pronoun in (2) should be easier to understand than the one in (1), because it has only one referent that matches the gender of the pronoun.

(1) Clinton confessed to Gore when he asked for the truth.

(2) Clinton confessed to Albright when she asked for the truth.

But what are the relative roles of accessibility and gender information in the on-line processes of pronoun interpretation? Research on this issue presents conflicting findings.

Some researchers have suggested that in contexts like (2), a referent may be assigned to a pronoun on the basis of gender information alone (e.g. Crawley, Stevenson & Kleinman, 1990; Ehrlich, 1980; see also Shillcock, 1982). The conclusions of these researchers suggest that gender is used first during processing, and other factors like referent accessibility may not be used at all, or at least not until a second stage. By contrast, theories like the Minimalist Hypothesis (McKoon & Ratcliff, 1992) give priority to accessibility factors. They suggest that pronoun interpretation occurs automatically only when there is a single, sufficiently accessible match for the referent; otherwise even cues like unambiguous gender will be used in a strategic, non-automatic fashion.

The results from on-line research on this question are also mixed. These studies have typically used a written probe paradigm, where participants read passages that are interrupted at various points by the presentation of a written probe word. Participants then rapidly judge whether they have seen the probe word before in the sentence. Somewhat surprisingly, many of these studies have failed to find clear and rapid effects of gender information on pronoun resolution. At one extreme, Greene, McKoon and Ratcliff (1992) found no evidence that participants identified the referents of gender-disambiguated pronouns at all. After reading a pronoun, participants were no faster to respond to a probe word related to the referent than to a probe word related to the non-referent, except in experiments with extremely slow text presentation, short texts, and comprehension questions that encouraged them to focus on the pronoun. Greene et al. (1992) concluded that participants do not automatically identify a unique referent unless accessibility information makes it easy. If it doesn't, even if a gender cue is present, participants must engage in special problem-solving strategies to identify a unique referent, as may occur when task demands lead them to focus on the pronoun. Authors of other probe studies have made similar claims about the limited use of gender information (Garnham, Oakhill & Cruttenden, 1992; Gernsbacher, 1989; McDonald & MacWhinney, 1995), and Vonk (1985) came to a similar conclusion with an eyetracking reading study.

By contrast, other researchers have used similar probe tasks and found rapid effects of pronoun resolution. In particular, Boland, Acker and Wagner (1998) found effects of referent facilitation for gender-disambiguated pronouns, even when the text was rapidly presented and participants were not encouraged to use special strategies. MacDonald and MacWhinney (1990), using a cross-modal probe paradigm, found that gender-disambiguated pronouns were processed more rapidly than ambiguous pronouns, although facilitation of the intended referent was inconsistent across their experiments. Other studies have also reported that gender cues facilitate pronoun resolution, but using methods that do not tap the initial processes of pronoun interpretation (e.g. Ehrlich, 1980; Hudson-D'Zmura & Tanenhaus, 1998). By contrast with these studies, which all investigate the use of semantic gender of referent entities, research on Romance languages has shown that pronoun resolution is influenced by the morphosyntactic gender of potential antecedents. For example, readers utilize the fact that 'la spia' (the spy) is feminine in Italian, regardless of the semantic gender of the referent (e.g. Cacciari, Carreiras & Cionini, 1997).

Thus, previous studies on the use of semantic gender have found inconsistent evidence that gender information is used rapidly to disambiguate the intended referent of a pronoun, suggesting fairly limited use of this cue by comprehenders. However, we raise several concerns about these previous studies, which could draw into question this conclusion. In particular, the majority of these studies required that on-going language comprehension be interrupted for a probe task, which may have hindered referential processing. In addition, all of these studies relied either directly or indirectly on the task of reading, which raises the question of how well referent resolution proceeds in the more common situation of uninterrupted spoken language comprehension.

Given these concerns and the mixed findings on this issue, we asked whether gender information has rapid on-line effects during pronoun interpretation, using the technique of monitoring participants' eye movements while they listened to spoken texts. This method, recently developed by Tanenhaus and colleagues (e.g. Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1996; see also Cooper, 1974), is less invasive and allows us to investigate speech processing rather than reading. In addition, this method provides moment-by-moment information about the referents that listeners are considering for both ambiguous and unambiguous pronouns, without interrupting the comprehension process.

1.2. Conditions of felicitous pronoun use

Before conducting the eyetracking studies, we established some of the conditions of felicitous pronoun use in an off-line written story-completion experiment. Seventeen participants read written sentences, e.g. 'The bride/groom embraced the maid of honor after the wedding', and added a natural continuation sentence. Each of the 12

Table 1
Percentage pronoun use in the story-completion experiment

	1st character (%)	2nd character (%)
Gender-different (unambiguous)	95	76
Gender-same (ambiguous)	95	37

experimental items contained two gender-specific noun phrases and occurred in two conditions: gender-same versus gender-different.

We analyzed their use of pronouns versus names for the first reference in their response, depending on whether they chose to refer to the 1st- or 2nd-mentioned character, and depending on whether a pronoun would be ambiguous or unambiguous by gender. The condition means (Table 1)[1] show that pronouns were used more than names in all conditions except when the pronoun was ambiguous and referred to the second (less accessible) character, resulting in a reliable interaction between these factors ($F1(1,15) = 7.4$, $P < 0.02$; $F2(1,11) = 10.9$, $P < 0.01$).

Given the interaction between gender and accessibility in our off-line data, we investigated how comprehenders use these factors in on-line pronoun comprehension. We report results from two experiments, both of which show that both gender and accessibility have rapid effects on pronoun resolution in a way that is consistent with the off-line data. The first experiment manipulates accessibility by order-of-mention; the second establishes a greater contrast in accessibility through a combination of order-of-mention, recency, and pronominalization.

2. Experiment 1

2.1. Participants/methods/materials

Sixteen participants from the University of Pennsylvania community spent approximately 45 min participating in the experiment in exchange for $6. All participants began learning English before age 5.

Participants' eye movements were recorded while they viewed a picture and listened to text describing the picture. Participants indicated whether the utterances were consistent with the picture by pressing 'yes' or 'no' buttons. The 16 stimulus items were always 'yes' answers; there were also 16 fillers (four 'yes', 12 'no') and three practice items (two 'yes', one 'no').

[1] Items were excluded from the analysis when participants did not begin their continuation with a reference to either character of interest ($n = 30$), did not begin with a new sentence ($n = 7$), did not make sense ($n = 1$) or interpreted 'waiter' as female ($n = 1$). Only direct references to the character were counted, excluding other forms of reference (possessive NPs ($n = 2$), demonstrative NPs ($n = 1$), and non-repeated NP anaphors ($n = 2$)). Data from one participant was excluded because during debriefing she said that she had been taught not to be 'repetitive', so she consciously used pronouns throughout. This left 149 data points and 16 participants in the analysis. There were four missing cells in the participant analysis and five missing cells in the item analysis; these were replaced with the mean score for that participant or item.

An ISCAN eyetracking visor was used (for details see Trueswell, Sekerina, Hill & Logrip, 1999). The scene image and the superimposed eye position, along with all auditory stimuli, were recorded once for each 33-ms frame of video tape, using a frame-accurate digital video recorder (a SONY DSR-30).

The pictures used known cartoon characters like Mickey Mouse and Donald Duck. Before each experiment, we confirmed that participants were familiar with the characters, and had them practice naming them. There were four variants of each item, manipulating (a) gender (same versus different) and (b) order-of-mention (1st versus 2nd). Each item was rotated through the four conditions in a Latin square design, generating four lists, each having forward and reverse orders. Examples are shown in Fig. 1.

Each text had four clauses, broken into two sentences, as in (3). The first clause mentioned the two characters; the second mentioned some other object in the picture. 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 half the items the 1st-mentioned character was on the right of the picture, in half it was on the left. A different set of subjects rated the plausibility of each character being the subject of the third predicate; there were no significant differences between the two characters ($t = 0.534$, $P = 0.60$).

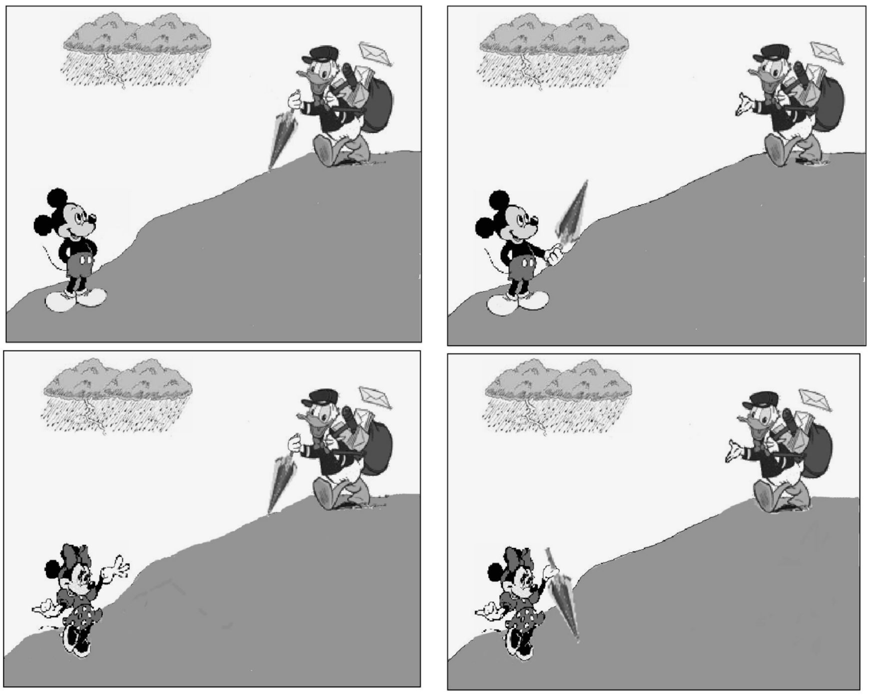

Fig. 1. Example stimulus pictures in each condition. The character carrying the umbrella is the referent of the pronoun.

(3) 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.

Texts were digitally recorded to computer. The same recording was used for all conditions of a given item, with the differing names and pronouns spliced in.[2] On each trial, a picture was presented simultaneously with the onset of the spoken text. The texts and participant responses were controlled by Psyscope 1.0.2 running on a Power Macintosh.[3]

2.2. Results

Participants' responses to the judgement task (whether the description matched the picture) indicated that they identified the correct pronoun referent in most cases. There was no difference among conditions in terms of either accuracy or response latency ($Ps > 0.2$),[4] but the means suggest that they were least accurate in the same/2nd-mention condition. The mean accuracy was 94% for different/1st-mention, 92% for different/2nd-mention, 92% for same/1st-mention, and 88% for same/2nd-mention.

We coded the video record of participants' eye movements, beginning at the onset of the pronoun. For each video frame, we identified where the participant was fixating: target (the referent of the pronoun), competitor (the other character), other (something else in or off the picture), or track loss.[5]

Fig. 2 presents the percentage of trials on which participants were looking at a given object, for each 33-ms frame. In the first three conditions (different/1st-mention, different/2nd-mention, and same/1st-mention), there were more looks to the pronoun referent than the other character immediately after the pronoun onset. This indicates that participants were using both accessibility and gender cue to guide their interpretation. In the fourth condition (same/2nd-mention), participants did not immediately converge on an interpretation of the pronoun referent, and instead looked equally at both target and competitor.

To test the reliability of target looks versus competitor looks, we split the data into four 200-ms segments. The first segment corresponded to the pronoun, which took an average of 200 ms. The second segment began at the onset of the verb, and the third and fourth began 200 and 400 ms after the verb, respectively. In each segment, trials with more than 33% track loss for that segment were excluded.

For each segment, we calculated the mean 'target advantage': the total amount of

[2] A colleague listened to each item and attempted to identify the original recording. She only succeeded in two of the 16 stimuli, indicating that the splicing was not noticeable.
[3] Due to an equipment malfunction, one participant reported half of his responses orally.
[4] Participants were not under any time pressure to respond, so variance in response latency was high.
[5] Track loss accounted for 8% of all the data.

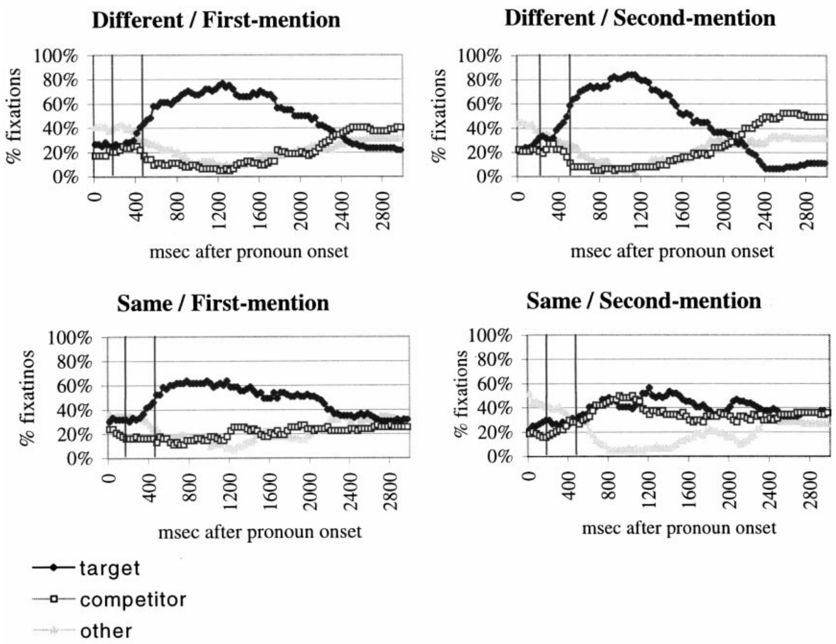

Fig. 2. Percentage of looks to each object in picture (on the *y*-axis) for each 33-ms frame on the video tape (on the *x*-axis). The graph begins at the onset of the pronoun and extends for 3 s. The lines mark the average onset and offset of the verb, which lasted an average of 302 ms.

time spent looking at the target in a given condition minus the total amount of time looking at the competitor.[6] ANOVAs were conducted on participant means (Table 2) and item means. In segments 1 and 2, no differences among conditions were reliable ($Fs < 2$). In the third segment, there was a reliable effect of gender ($F1(1, 15) = 6.0$, $P < 0.05$; $F2(1, 15) = 4.6$, $P < 0.05$) and an interaction between gender and referent ($F1(1, 15) = 5.4$, $P < 0.05$; $F2(1, 15) = 15.4$, $P < 0.002$). In the fourth segment, there was also a reliable effect of gender ($F1(1, 15) = 9.1$, $P < 0.01$; $F2(1, 15) = 9.2$, $P < 0.01$), and an interaction between gender and referent ($F1(1, 15) = 9.3$, $P < 0.01$; $F2(1, 15) = 12.8$, $P < 0.01$). There was also a marginal effect of referent in the items analysis ($F1(1, 15) = 2.5$, $P = 0.13$; $F2(1, 15) = 3.4$, $P = 0.08$).

To compare target and competitor looks in the segments with significant interactions (segments 3 and 4), *t*-tests were performed on the mean arcsine-transformed target and competitor looking times (over both participants and items). All segments showed more looks to the target ($ts < -4$, $Ps < 0.005$), except for the same-gender/2nd-referent condition ($|t|s < 1$, $Ps > 1$). In sum, participants

[6] The percentage of looks to target and looks to competitor for each participant and item were first transformed by arcsine($2P - 1$) in order to take into account that these values were bounded at 0 and 1.

Table 2
Target advantage in Experiment 1: difference in amount of time spent looking at target versus competitor (% target looks − % competitor looks)[a]

	Segment 1 (pronoun onset to verb onset (about 200 ms)) (no effects)	Segment 2 (next 200 ms) (no effects)	Segment 3 (next 200 ms) (interaction)	Segment 4 (next 200 ms) (interaction)
Different-gender/1st-mention	9	1	36	54
Different-gender/2nd-mention	4	13	52	70
Same-gender/1st-mention	12	15	44	57
Same-gender/2nd-mention	11	9	3	2

[a] Values are given as percentages.

looked at the target reliably more than the competitor in all conditions but same-gender/2nd-referent.

3. Experiment 2

Experiment 1 showed that listeners are able to rapidly converge on the intended referent based on order-of-mention and gender cues. Only when both of these cues were unhelpful (in the same-gender/2nd-mention condition) did listeners show difficulty identifying the pronoun referent.

However, the 2nd-mentioned character was still relatively accessible. Participants usually responded that the text matched the picture in this condition, which indicated that they had succeeded in eventually identifying the intended referent. Furthermore, participants in the off-line study used 37% pronouns for reference to the 2nd-mentioned character in the same-gender condition. What happens when the 2nd-mentioned character is truly inaccessible?

Our second experiment investigated the use of gender under conditions with a greater contrast in referent accessibility.

3.1. Participants/methods/materials

Sixteen students from the University of Pennsylvania spent approximately 45 min participating in the experiment for course credit. All participants were monolingual English speakers.

The same methods and visual stimuli were used. The texts were changed to increase accessibility of the 1st-mentioned character by adding a pronominal reference to this character, as in (4).

(4) Donald is bringing some mail to {Mickey/Minnie}.
He's sauntering down the hill,
while a violent storm is beginning.
He's/She's carrying an umbrella,
and it looks like they're both going to need it.

We predict from this change that the 2nd-mentioned character will be so inaccessible that it will not even be considered in the same-gender conditions. If so, eye movements in this condition will show that participants think the referent is the competitor (the 1st-mentioned character), and their responses will indicate that they often think the text and picture do not match (e.g. because Donald, not Mickey, should be holding the umbrella). Experiment 2 also strongly tests the effectiveness of gender in the different-gender/2nd-mention condition. If subjects easily identify the referent in this condition (e.g. Minnie), it means gender is a highly effective cue, even in the face of other cues that decrease the referent's accessibility.

3.2. Results

3.2.1. Effects of gender and accessibility

The results replicated the findings from Experiment 1, with differences going in the expected direction. In this experiment, pronominal reference to the second character in the same-gender condition was highly infelicitous, and subjects often garden-pathed in the same-gender/2nd-mention condition. This is evident in the response data, which show that subjects rejected the text 41% of the time in the same-gender/2nd-mention condition (in contrast to 14% of the time in different/1st-mention, 5% in different/2nd-mention and 5% in same/1st-mention). ANOVAs showed reliable main effects of gender and referent, and an interaction between the two ($Fs > 6$, $Ps < 0.05$, except in the items analysis where the effect of referent was marginal: $F2 = 3.8$, $P = 0.07$).

Eye movement data (Fig. 3) also replicated the results of Experiment 1, in that early target identification occurred when either gender or accessibility information was present. Participants rapidly converged on the target even in the different-gender/2nd-mention condition, revealing the effectiveness of the gender cue. However, the 2nd-mentioned character was so inaccessible, eye movements in the same-gender/2nd-mention condition revealed an early competitor advantage.

We tested the reliability of these results using the same procedure as in Experiment 1. The participant means are shown in Table 3. In the first segment there were no differences among conditions. In the second segment there was an interaction between gender and referent ($F1(1,15) = 8.2$, $P < 0.05$; $F2(1,15) = 5.0$, $P < 0.05$), showing up one segment earlier than in the previous experiment. In the third segment there was a main effect of gender ($F1(1,15) = 10.5$, $P < 0.01$; $F2(1,15) = 11.7$, $P < 0.005$) and an interaction between gender and referent ($F1(1,15) = 33.0$, $P < 0.001$; $F2(1,15) = 26.3$, $P < 0.001$). In the fourth segment there were main effects of gender ($F1(1,15) = 86.7$, $P < 0.001$; $F2(1,15) = 43.0$, $P < 0.001$), referent ($F1(1,15) = 16.0$, $P < 0.001$; $F2(1,15) = 11.4$, $P < 0.005$),

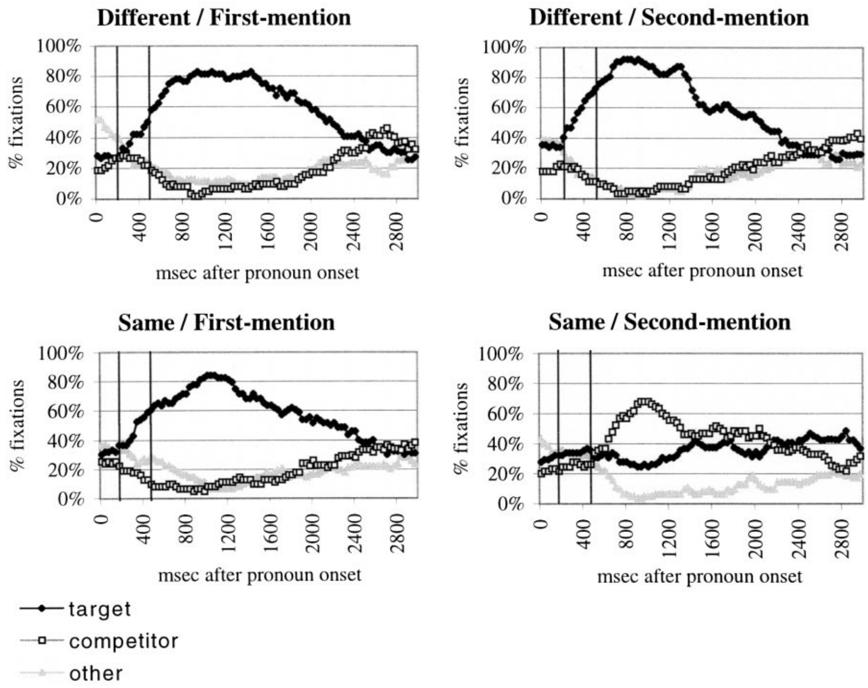

Fig. 3. Percentage of looks to each object in picture (on the y-axis) for each 33-ms frame on the video tape (on the x-axis). The graph begins at the onset of the pronoun and extends for 3 s. The lines mark the average onset and offset of the verb, which lasted an average of 287 ms.

and an interaction between the two ($F1(1, 15) = 48.7$, $P < 0.001$; $F2(1, 15) = 54.6$, $P < 0.001$). All other $Fs < 3$, $Ps > 0.1$.

We again performed t-tests to compare target and competitor looks in the third and fourth segments, and found that participants were looking at the target significantly more than the competitor in the first three conditions ($ts < -3$, $Ps < 0.01$). By contrast, in the same-gender/2nd-mention condition, there was no difference in the third segment ($|t|s < 1$, $Ps > 1$), and in the fourth segment subjects looked more at the competitor than the target (although the test was marginal in the participants analysis; $t1 = 3$, $P < 0.01$; $t2 = 1.9$, $P = 0.08$).

3.2.2. Point of disambiguation

One question that arises from these results concerns the data from the same-gender/2nd-mention condition. Since neither gender nor accessibility aids pronoun comprehension in this condition, we might expect participants to experience a strong garden path, and look initially to the 1st-mentioned character, the competitor. However, in Experiment 1 they looked equally to target and competitor, and in Experiment 2 the initial competitor advantage appeared smaller than the target advantage in other conditions. Why were participants having any success at all in this condition?

Table 3
Target advantage in Experiment 2: difference in amount of time spent looking at target versus competitor (% target looks − % competitor looks)[a]

	Segment 1 (pronoun onset to verb onset (about 216 ms)) (no effects)	Segment 2 (next 200 ms) (interaction)	Segment 3 (next 200 ms) (interaction)	Segment 4 (next 200 ms) (interaction)
Different-gender/1st-mention	8	14	33	66
Different-gender/2nd-mention	14	33	68	83
Same-gender/1st-mention	8	35	49	66
Same-gender/2nd-mention	9	4	3	− 29

[a] Values are given as percentages.

In both experiments, we hypothesize that the reason for their partial success was that some items provided early disambiguating information. In some items, like ‘She is singing along with the music’, the disambiguating information comes at the verb, although for some it is only partial and probabilistic. In these items, participants were able to use verb information to inform their initial interpretations of the pronoun. In other items, like ‘He is wearing a hat’, the disambiguating information comes late, at ‘hat’. Half the items were categorized as ‘early disambiguation’ items, half as ‘late’.

A comparison of early versus late disambiguating items in the same-gender/2nd-mention condition supported this hypothesis in both experiments (Fig. 4). In Experiment 1, there was an early target advantage for early disambiguation items, and an early competitor advantage for late disambiguation items. In Experiment 2, the 2nd-mentioned character was so inaccessible that participants still had partial difficulty in the early disambiguation items, and the late disambiguation items showed a very large competitor advantage.[7]

Importantly, there was no point of disambiguation effect in the other three conditions, where accessibility and gender information supported early identification of the target: participants looked to the correct referent (the target) and continued to look there regardless of the point of disambiguation.

4. General discussion

The main finding in both experiments was that both gender and accessibility affect

[7] ANOVAs were performed to compare ‘early’ and ‘late’ conditions in each experiment for segments 3, 4, and a subsequent 5th 200-ms segment. In both experiments, the test was reliable or marginally reliable for segment 5 (Experiment 1: $F2(1,14) = 3.9$, $P = 0.068$; Experiment 2: $F1(1,15) = 8.3$, $P < 0.02$; $F2(1,14) = 5.2$, $P < 0.05$). Disambiguation point was not included as a planned factor in Experiment 1, so participant analyses could not be performed due to missing cells. Although the tests did not reach significance for segments 3 and 4 ($Fs < 3$, $Ps > 0.1$), this is likely to be due to the low number of items in each sample, and the fact that the point of disambiguation is not perfectly aligned in each group.

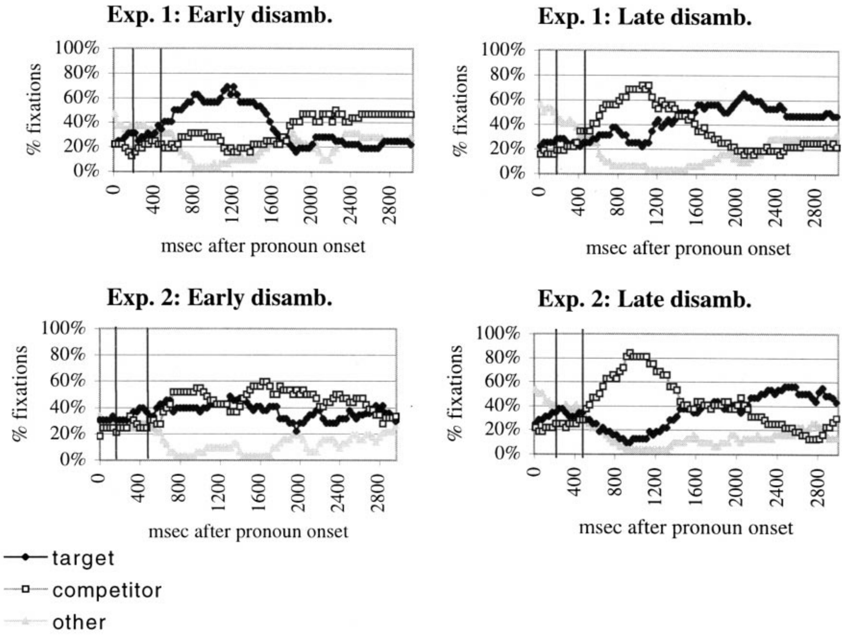

Fig. 4. Probability graphs for the same-gender/2nd-mention condition in Experiments 1 and 2, split by disambiguation point. The left edge of the graph corresponds to the pronoun onset, and the lines correspond to the verb onset and offset.

the initial processes of pronoun resolution. When either factor signaled the correct pronoun referent, participants' eye movements indicated that they began considering the target referent more than the competitor referent soon after the offset of the pronoun. When neither gender nor accessibility was sufficient to resolve the pronoun, participants did not rapidly converge on an interpretation of the pronoun. These findings mirrored the off-line measure of pronoun plausibility, where pronouns were preferred in all cases except when gender and accessibility were not available to help pronoun interpretation.

These experiments also provide important evidence about the time course of pronoun resolution. Using spoken stimuli and a relatively natural task, we found evidence of pronoun resolution beginning between 0 and 200 ms after the offset of the pronoun. This latency is comparable with the identification of more explicit referential forms in other eyetracking studies (e.g. Allopenna, Magnuson & Tanenhaus, 1998). Thus, the results show rapid identification of the pronoun referent, even though participants were not instructed to respond quickly.

Our results further show that gender and accessibility information influence referent consideration at the same moment. This finding runs counter to two types of claims that have been made in the literature. First, it shows that gender information is not used at a stage prior to accessibility information (cf. Crawley et al., 1990; Ehrlich, 1980). Second, it doesn't support the claim that gender information is only used during special, strategic, or later processing (cf. e.g. Garnham et al., 1992; Gernsbacher, 1989; Greene et al., 1992). Our study employed a natural task, had no comprehension questions, and used a normal rate of presentation. Therefore, it is difficult to conclude that participants used strategies that are not present in the normal course of language use. Moreover, participants rapidly used gender to determine the referent even when other cues drastically decreased the accessibility of the intended referent, as found in the different-gender/2nd-mention condition of Experiment 2.

Why, then, did some previous studies fail to find an effect of gender? The written probe methodology has several characteristics that may have made it difficult to find an effect of gender. First, it interrupts comprehension processes, placing a memory load on the reader, and draws attention toward the surface representation of the text, rather than the meaning. These facts might hinder the construction of rich discourse representations, especially since the texts used unfamiliar, unelaborated characters that changed for every item. The lack of context is also likely to decrease participants' motivation for fully understanding the texts. By contrast, our study used spoken data, familiar characters, and a visual context – characteristics that are common to everyday language use.

The results in our study support a dynamic model of language processing, where multiple sources of information are used probabilistically to guide referential processing. Arnold (1998) suggested that referent accessibility is influenced by the comprehender's estimation of the probability that a given referent will be mentioned again in the discourse, and that this estimation is built probabilistically on the basis of multiple cues, such as order-of-mention and recency. Our results are consistent with this account, and suggest that gender information is yet another constraint that contributes probabilistically to pronoun resolution. In this study, gender and accessibility played nearly equal roles, but in other contexts one cue may be stronger or more informative than the other, in which case it would have a larger impact. However, the results here show that neither gender nor accessibility are ignored during the initial processes of pronoun resolution. This suggests that it is not appropriate to ask whether gender or accessibility is used first, and instead we should aim to discover their relative contributions under different discourse conditions.

Acknowledgements

This work was partially funded by National Institute for Health Grant 1-R01-HD37507-01 and by National Science Foundation Grant SBR-96-16833. We thank the Institute for Research in Cognitive Science at the University of Pennsylvania (NSF-STC Cooperative Agreement number SBR-89-20230) for the use of their eyetracking facilities. Many thanks to Jared Novick for his assistance with some of the eye movement analyses.

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.

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

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.

Cacciari, C., Carreiras, M., & Cionini, C. B. (1997). When words have two genders: anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, *37*, 517–532.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6* (1), 84–107.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, *194*, 245–265.

Ehrlich, K. (1980). Comprehension of pronouns. *Journal of Experimental Psychology*, *32*, 247–255.

Garnham, A., Oakhill, J., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes*, *73* (4), 231–255.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*, 99–156.

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

Hudson-D'Zmura, S., & Tanenhaus, M. K. (1998). Assigning antecedents to ambiguous pronouns: the role of the center of attention as the default assignment. In M. Walker, A. Joshi, & E. Prince, *Centering theory in discourse* (pp. 199–226). Oxford: Oxford University Press.

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

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.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *993*, 440–466.

Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech*, *25* (Part 4), 385–401.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: evidence for visually mediated incremental interpretation. In T. Inui, & J. L. McClelland, *Attention and performance XVI: information integration in perception and communication* (pp. 457–478). Cambridge, MA: MIT Publishing.

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

Vonk, W. (1985). On the purpose of reading and the immediacy of processing pronouns. In R. Groner, G. W. McConkie, & C. Menz, *Eye movements and human information processing*. Amsterdam: Elsevier.