Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 4, 874–887

0278-7393/11/$12.00 DOI: 10.1037/a0022932

Rules of Engagement: Incomplete and Complete Pronoun Resolution

Jessica Love and Gail McKoon
Ohio State University

Research on shallow processing suggests that readers sometimes encode only a superficial representation of a text and fail to make use of all available information. Greene, McKoon, and Ratcliff (1992) extended this work to pronouns, finding evidence that readers sometimes fail to automatically identify referents even when these are unambiguous. In this paper we revisit those findings. In 11 recognition probe, priming, and self-report experiments, we manipulated Greene et al.'s stories to discover under what circumstances a pronoun's referent is automatically understood. We lengthened the stories from 4 to 8 lines. This simple manipulation led to automatic and correct resolution, which we attribute to readers' increased engagement with the stories. We found evidence of resolution even when the additional text did not mention the pronoun's referent. In addition, our results suggest that the pronoun temporarily boosts the referent's accessibility, an advantage that disappears by the end of the next sentence. Finally, we present evidence from memory experiments that supports complete pronoun resolution for the longer but not the shorter stories.

Keywords: pronouns, reading, engagement, recognition memory

Consider this excerpt from Flannery O'Connor's short story "A Good Man Is Hard to Find":

> The children's mother put a dime in the machine and played "The Tennessee Waltz" and the grandmother said that tune always made her want to dance. (O'Connor, 2001, p. 1138)

Is it the case that "The Tennessee Waltz" makes the grandmother want to dance? Or perhaps it is the children's mother who enjoys dancing to the song? Pronominal ambiguity is widespread, both in spoken and written English, and many researchers are understandably interested in the textual factors that affect which of two or more pronominal referents is chosen as correct. However, we believe that the question of whether a pronoun is resolved at all deserves to be reexamined.

In this article, we first provide a context for our exploration of incomplete pronoun resolution, paying special attention to recent research on the resolution of nominal anaphora. We then report findings from recognition probe, priming, and self-report experiments designed to identify when pronouns are resolved completely and to what effect. In Experiments 1 and 3, we replicated the findings of Greene, McKoon, and Ratcliff (1992) that readers do not automatically resolve an unambiguous pronoun when it is presented, at a normal reading speed, with short stories such as the following:

> Rita and Walter were writing an article for a magazine. They had to get it done before next Tuesday. Rita edited the section that Walter had written and then she smoked a cigarette to relax.

We used a recognition probe task, where test words appear at varied points in a story and participants respond "Yes" if they recognize a word as being from the story they are currently reading. In Experiments 1 and 3, we found that the relative accessibility of the referent, as measured by this task, does not increase with respect to the nonreferent after the pronoun's appearance. In Experiments 2 and 4, we lengthened the stories from four to eight sentences. This simple manipulation led to data consistent with automatic resolution. The presence of additional text allowed the relative accessibility of the referent to increase after the pronoun, a finding we attribute to readers' increased engagement with the text.

In the remainder of the experiments, we explored the nature of the increased accessibility that, we argue, signals anaphoric resolution. Experiment 5 extended our results to lengthened stories that contained no additional mention of the two characters. In Experiments 1 through 5, the referent of the pronoun was the subject of the preceding clause ("Rita"). In Experiment 6, the referent was the object of the preceding clause ("Walter"). Experiment 7 confirmed that it was the pronoun's effect on the referent, not the nonreferent, that allowed resolution to occur. Experiment 8 traced the time course of the referent's accessibility across sentences. In Experiments 9 and 10, we used an offline priming task to examine the effects of pronominal resolution on the later representations of the stories in memory. Finally, in Experiment 11, we used a questionnaire to ask participants directly whether the longer stories were more engaging than the shorter ones. Data from all of these experiments are consistent with automatic resolution for the longer but not the shorter experimental stories.

Background

Pronoun resolution can be understood in terms of a discourse model in which discourse events and the entities involved in the events are represented (Greene et al., 1992; Grosz, 1981; Grosz, Joshi, & Weinstein, 1983; Grosz & Sidner, 1986; Rigalleau, Caplan, & Baudiffier, 2004; Sidner, 1983a, 1983b; Stewart, Holler, &

This article was published Online First April 11, 2011.

Jessica Love and Gail McKoon, Department of Psychology, Ohio State University.

Preparation of this article was supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC01240 to Gail McKoon.

Correspondence concerning this article should be addressed to Jessica Love, Department of Psychology, 291 Psychology Building, Ohio State University, Columbus, OH 43201. E-mail: love.131@osu.edu

Kidd, 2007; Webber, 1983). In our view, discourse entities have a degree of accessibility that changes as the local discourse representation changes. This accessibility is determined by the syntactic structure of the discourse, the semantic relationships among discourse entities, and general knowledge already familiar to the reader. The accessibility of a discourse entity is measured relative to the past and current accessibilities of other entities in the text. At any point in the text, the entities that are most accessible are what is in focus (i.e., what the discourse is "about").

We understand the features of a pronoun (i.e., animacy, gender, and number) to be matched automatically, passively, and in parallel against the semantic features of other entities in the discourse and all other entities in memory (Gillund & Shiffrin, 1984; Hintzman, 1984; Murdock, 1982). Greene et al. (1992) proposed that the referent of a pronoun is automatically understood only if that referent is "sufficiently more highly accessible in the comprehender's discourse model relative to the pronoun as a memory cue than all other discourse entities" (p. 267). This makes available the information necessary for resolution, where the referent must be integrated into a reader's representation of the text in memory (i.e., information about the pronoun must be linked to information about the referent). This approach to pronoun resolution, then, predicts that if no discourse entity matches sufficiently, or if more than one entity matches sufficiently, this process of instantiation does not occur and the pronoun's referent is not fully integrated.

Much of the time, pronouns are used when there is only a single possible referent. Pronoun resolution often is automatic and complete, and it is unsurprising that many researchers design and interpret experiments under this premise. Much of the ongoing research on anaphor resolution, thus, focuses on how language comprehenders use sentence position or pragmatic clues to choose among possible referents (Arnold, 2001; Crawley, Stevenson, & Kleinman, 1990; Marslen-Wilson, Tyler, & Koster, 1993; Rohde, Kehler, & Elman, 2006; Smyth, 1994; Stevenson, Crawley, & Kleinman, 1994). Although this research has contributed to our understanding of pronoun resolution considerably, it often assumes something we believe to be incorrect: that a single discourse referent is always understood.

Both past and present research provide ample evidence that readers often engage in processing that yields discourse representations that are underrealized and that computes and infers only information that is easily and automatically available, unless task requirements demand more (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Erickson & Mattson, 1981; Ferreira, Bailey, & Ferraro, 2002; McKoon & Ratcliff, 1992; Sanford, 2002; Sanford & Sturt, 2002). McKoon and Ratcliff's minimalist approach, currently embodied in resonance theory (e.g., Gerrig & O'Brien, 2005; Myers & O'Brien, 1998), argues for strict constraints on the kinds of inferences readers make automatically, arguing that readers often proceed with minimally complex textual representations. For example, McKoon and Ratcliff (1986) had participants read passages such as "The director and the cameraman were ready to shoot closeups when suddenly the actress fell from the 14th story." They found that, rather than making the more explicit inference that the actress died, participants appeared to encode simply that something bad had happened.

It is in the context of readers' tendencies toward shallow, underspecified discourse representations that the question of whether an anaphor is resolved comes to attention. Previous research addresses this question for one type of anaphora, noun anaphora, and suggests that the referents of nominal anaphors are identified correctly and quickly only in ideal circumstances, when they have no plausible competition (Dell, McKoon, & Ratcliff, 1983; Levine, Guzmán, & Klin, 2000; McKoon & Ratcliff, 1980a). Using a probe recognition task, McKoon and Ratcliff (1980a) presented participants with stories such as the following:

> A burglar surveyed the garage set back from the street. Several milk bottles were piled at the curb. The banker and his wife were away on vacation. The criminal slipped away from the streetlamp.

They found that the appropriate referent *burglar,* as well as any discourse entity that is propositionally related to it (e.g., *garage*), becomes more accessible after the mention of the noun anaphor *criminal.* That this increase in accessibility, measured relative to the accessibilities of other discourse entities, started to appear at just 250 ms after the presentation of the anaphor is interpreted as evidence that noun anaphors are processed automatically.

Recently, however, Levine et al. (2000) showed participants stories containing a referent *tart,* followed by an elaborately described, more predictable competitor *cake.* Their results suggest that participants did not resolve the anaphor *dessert,* which was mentioned later in the short text. Only when relevant parts of the text were highlighted with asterisks and participants were instructed to pay special attention to these highlighted parts did evidence support the correct resolution of the anaphor (see also Klin, Guzmán, Weingartner, & Ralano, 2006).

Levine et al. (2000) pointed out that noun anaphora may be processed differently than pronouns. A noun such as *dessert* contains a great deal more information than a pronoun, which can specify at most animacy, gender, number, and case in English. Thus, although a story may still be coherent with a slightly underspecified representation of dessert, it may lose coherence if the much sparser representations of pronouns are not resolved, making resolution more important for pronouns than noun anaphors.

Nonetheless, Greene et al. (1992) showed that pronouns, too, can be left unresolved. Even when a pronoun's referent is available–even, in fact, when it is unambiguous–there are still situations in which readers do not automatically link the information about the pronoun to information about the referent in memory. In Green et al.'s study, participants were presented with four-line stories, such as the story about Rita and Walter given previously (see also Table 1). Each story contained exactly one female and one male character. The names of both characters were mentioned in the first line, as well as in the third line. The fourth line of the story contained a pronoun that referred unambiguously to one of the two characters (e.g., "and then she smoked a cigarette to relax"). Participants were presented with the stories at 250 ms per word, a normal reading pace (Just & Carpenter, 1987), and were asked to respond "Yes" if a test word, *Rita* or *Walter,* had appeared in the story. Participants were tested either immediately before the pronoun *she* in the final line of the story or directly after the story's completion. There was no difference in the pattern of response times to characters that did or did not match the gender of the pronoun.

Greene et al. (1992) interpreted this result in terms of the passive, automatic matching process described above. In the environment of *she,* there are at least three highly accessible entities:

Table 1
Sample Stimuli for Experiments 1, 2, 3, and 4

Story 1	
Exp. 1–4	Rita and Walter were writing an article for a magazine.
	They had to get it done before next Tuesday.
Exp. 2 and 4	Rita didn't trust Walter to get the facts right.
	Once, he'd written a piece about aliens landing in Chicago.
	"I'm going to get dragged down with you," Rita said at the time.
	However, neither of them had been fired.
Exp. 1–4	Rita edited the section Walter had written
Exp. 1 and 2	and then [TEST] she smoked a cigarette to relax. [TEST]
Exp. 3 and 4	and then [TEST] she smoked a cigarette [TEST] to relax.
Test words	Rita (referent), Walter (nonreferent)
Story 2	
Exp. 1–4	Tracy and Arthur had been smuggling drugs for years.
	They were quite proficient with a well-practiced routine.
Exp. 2 and 4	Tracy needed the money to support a nasty drug habit.
	She was always nervous right before a run.
	"Are you going to back out now?" asked Arthur.
	He had already risked so much.
Exp. 1–4	Tracy got the drugs from Arthur to hide in a stewardess bag,
Exp. 1 and 2	and then [TEST] she carried the bag past customs. [TEST]
Exp. 3 and 4	and then [TEST] she carried the bag [TEST] past customs.
Test words:	Tracy (referent), Arthur (nonreferent)

"Rita," "Walter," and "cigarette." Among the features of these entities are those that match the features of *she*: feminine and singular. In this situation, Greene et al. argued, comprehension could proceed without actually resolving *she,* that is, without understanding it as referring to Rita and without linking information pertaining to the *she* (i.e., smoked a cigarette) to the representation of Rita in memory. A speedup for the correct referent *Rita* relative to that of the nonreferent *Walter* appeared only when the stories were presented much more slowly, at 500 ms per word, and participants were given comprehension questions that explicitly probed their understanding of the pronoun (see also Gernsbacher, 1989).

In their discussion of nominal anaphora, Levine et al. (2000) argued that the probability of identifying a correct referent should be a function of two factors: the degree of accessibility of the referent and the extent to which resolution is necessary to create a coherent discourse representation. If there is no referent that is more accessible than any other and the reader's implicit "standard of coherence" (van den Broek, Risden, & Husebye-Hartmann, 1995) is met without it, readers will continue undeterred.

We agree with Levine et al. (2000) that the probability of identifying a referent should be a function of two factors, one of them being the degree of accessibility of the referent. However, we expand the standard of coherence to include not only the extent to which resolution is necessary to create a coherent discourse representation but also the reader's engagement in the text. Interest and engagement in a text have been shown to influence how well a text is remembered and understood in several ways (Schiefele, 1991; Schraw, Bruning, & Svoboda, 1995). Education researchers have long known that both interest in a topic and intrinsic motivation to succeed are positively correlated with various indicators of academic learning, such as grades and scores on achievement tests (Schiefele & Schreyer, 1994). The correlation also holds for text learning. Participants' interest ratings for a text topic are positively correlated with performance on later tests of comprehension, free recall, and the ability to apply knowledge from one text to another. Schiefele and colleagues also found correlations between participants' levels of intrinsic motivation, as manipulated experimentally, and later test performance (Schiefele, 1996; Schiefele & Schreyer, 1994). Other researchers have shown a relationship between the amount of self-reported "transportation" from the real world to a story world (perhaps the highest form of narrative engagement) and agreement with a tacit belief advocated by the story (Green & Brock, 2000). The more the reader connects with the events described in a narrative, the more the reader will allow the narrative to shape his or her real-world beliefs. These findings indicate that reader engagement is powerful in influencing learning and belief. Thus, it is not unreasonable to expect that engagement can influence the degree to which textual inferences, including the inferences involved in anaphoric resolution, are generated.

Consider an excerpt from Raymond Carver's short story "A Small, Good Thing." A young boy, Scotty, has just awoken from a coma, and his parents Howard and Ann are at his side.

> They leaned over the bed. Howard took the child's hand in his hands and began to pat and squeeze the hand. Ann bent over the boy and kissed his forehead again and again. She put her hands on either side of his face. (Carver, 1989, p. 396)

Even though the pronoun *she* in the final sentence is unambiguous, casual readers may not fully update their understanding of the story to reflect that it was Ann—the same Ann who kisses Scotty's forehead—holding Scotty's face. A more engaged reader, perhaps one who has read the entire story and can take excitement in Scotty's awakening, might automatically make this connection.

In the experiments in this article, we revisited the findings of Greene et al. (1992), because, although we have argued that there are instances in which pronouns are not automatically resolved, we know there are many instances in which they are, at normal reading speed and without task-specific instructions. Our question was whether we could manipulate the stimuli used by Greene et al. such that pronouns are fully and automatically understood (i.e., such that a pronoun's referent is identified and information attributed to the pronoun is attributed to the referent).

In Experiments 1 and 2, we investigated the effects of reader engagement on pronoun resolution. In Experiment 1, we replicated Greene et al. (1992), finding no evidence of automatic, unambiguous pronoun resolution when participants were presented with four-line stories. In Experiment 2, we attempted to increase readers' engagement in the stories by simply adding four lines of text to the stimuli. We wrote the additional lines to contain as much interesting information as possible. It was our hope that making the stories longer and more interesting would increase readers' engagement in the stories, perhaps resulting in a higher standard of coherence and thus increasing the likelihood of pronoun resolution.

Experiments 1 and 2

These experiments used a probe recognition task to determine the relative levels of accessibility of referents and nonreferents before and after the presentation of a pronoun. In Experiment 1, participants were presented with 28 four-line stories, a subset of the stimuli used by Greene et al. (1992). Each story contained two characters, one female and one male, and the pronoun mentioned in the final line referred unambiguously to one of the two (see Table 1). In Experiment 2, participants were presented with modified versions of the same stories. Four lines were added between the first two lines and the last two lines of a story. For the example given previously, those four lines were the following:

> Rita didn't trust Walter to get the facts right. Once, he'd written a piece about aliens landing in Chicago. "I'm going to get dragged down with you," Rita said at the time. However, neither of them had been fired.

For both experiments, participants were asked to respond "Yes" to a test word if they had seen the word anywhere in the story and "No" if they had not. The test words for the stories of interest were the names of the two characters in the story.

Method

Materials. The 28 experimental stories used in Experiment 1 were three-sentence, four-line stories, a subset of those used by Greene et al. (1992). The stories each contained a male and a female character. The first-mentioned character of the first and third sentences was male for half of the stories and female for the other half. The pronoun in the second clause of the third sentence was always of the same gender as the first-mentioned character. The test words for the stories were the two character names. The positions for the test words were immediately before the pronoun and at the end of the sentence containing the pronoun. The number of words between the pronoun and the end of the sentence averaged 5.3 ($SD = 0.55$).

For Experiment 2, four additional lines were inserted into the middle of each story such that the first two lines of the story and the last two lines were identical to those in the original version. The referent character was mentioned an average of 2.9 times in the added lines, and the nonreferent character was mentioned an average of 3.0 times. The test words and test positions were the same as those in Experiment 1.

In both experiments, there were also 32 filler stories used to provide different kinds of test words and different testing locations from those in the experimental stories. Filler stories were written to match the experimental stories in length at four lines long (Experiment 1) or eight lines long (Experiment 2). The shorter fillers contained a total of 32 test words, 7 positive and 25 negative. Of the 32 filler test words, 14 were names (0 positive, 14 negative) and 18 were nonnames, usually nouns (7 positive, 11 negative). The eight-line fillers contained a total of 54 test words, 13 positive and 41 negative. Of the 54 filler test words, 20 were names (4 positive, 16 negative) and 34 were nonnames, usually nouns (9 positive, 25 negative). When test words were included in the experimental stories, the correct response was "Yes" for 58% of the test words in Experiment 1 and for 50% of the test words in Experiment 2.

A true or false comprehension statement was written for each of the experimental and filler stories. Half of these statements were written to be true according to the passage, and the other half were written to be false. None of these test statements required knowing the referent of a pronoun in order to make a correct response.

Procedure and participants. For all of the experiments reported in this article, all of the stories and test items were presented on a PC screen and responses were collected on the PC keyboard. Each participant completed one session of about forty minutes. The participants for all the experiments reported in this article were Ohio State University undergraduates taking part in the experiments for credit in an introductory psychology course. In each experiment, there were 32 participants.

The experiments began with 16 lexical decision test items, included to give participants practice with the response keys on the PC keyboard. After this practice, there were four filler stories, and then the remainder of the stories—28 experimental stories and 28 filler stories—were distributed randomly into 14 blocks such that each block contained two experimental and two filler stories. A different random order of presentation of materials was used for every second participant.

Each story began with the instruction to press the space bar on the keyboard to initiate the text. When the space bar was pressed, the story appeared one word at a time. Each word was displayed for 250 ms, then the next word was displayed for 250 ms, and so on until a complete line of the story appeared across the screen. Then the whole line disappeared and the next line was displayed in the same manner. When a test word was presented, the current line of text was erased and the test word appeared where the next word would have been. The letters of the test word were all in uppercase (unlike the words of the story), and two asterisks were displayed immediately to its right. The test word remained on the screen until a response key was pressed (*?/* for "Yes, the word appeared in the story" and *Z* for "No, the word did not appear in the story"). After the response and a pause of 100 ms, the story continued, unless the response was an error or the response was too slow. If the response was an error, the word *ERROR* was displayed for 900 ms before the story continued. If the response was slower than 1,500 ms, the message *TOO SLOW* was displayed for 900 ms. Participants were instructed to respond to the test words as quickly and accurately as possible. After each story, a true or false comprehension statement was presented. It remained on the screen until a response key was pressed (*?/* for "True" and *Z* for "False").

Design. For both experiments, there were two variables: the two test positions were crossed with the two test words. Test positions and test words were counterbalanced for participants and items in a Latin-square design.

Results and Discussion

Mean response times and accuracy values for the experiments are presented in Table 2. Experiment 1 replicated the findings of Greene et al. (1992). With short, four-line stories, there was no significant speedup from before the pronoun to the end of the sentence for a referent test word compared to a nonreferent test word. The critical interaction between the two test words and the two test positions was not significant, $F_1(1, 31) = 2.5$, $p = .12$, and $F_2(1, 24) = 1.8$, $p = .19$. The main effects of test word and test position were not significant, $F_1(1, 31) = 3.0$, $p = .09$, and $F_2(1,$

Table 2
Results for Experiments 1 and 2: Mean Response Times (RTs) and Probabilities Correct

Experiment	Test word	Test position	Mean RTs (probabilities correct)
1 (short versions)	Referent	Before pronoun	837 ms (.93)
		After pronoun	858 ms (.91)
	Nonreferent	Before pronoun	830 ms (.96)
		After pronoun	819 ms (.97)
2 (long versions)	Referent	Before pronoun	882 ms (.97)
		After pronoun	879 ms (.97)
	Nonreferent	Before pronoun	847 ms (.98)
		After pronoun	902 ms (.96)

24) = 1.8, $p = .19$ (Fs < 1.0, ps >.3, respectively). The 95% confidence interval on the response time means (see Loftus & Masson, 1994, for within-subject designs) was 20 ms.

Participants responded differently to the test words in Experiment 2. With the long stories, response times for the nonreferent slowed from before the pronoun to the end of the pronoun sentence by 55 ms, but response times for the referent did not (they were 3 ms faster). This interaction was significant, $F_1(1, 31) = 7.1$, $p < .05$, and $F_2(1, 24) = 6.1$, $p < .05$. Therefore, unlike Experiment 1, it appears that the presence of a pronoun did boost the relative accessibility of the referent. For the other effects, the main effect of test word was significant for items but not participants, $F_1(1, 31) = 2.8$, $p = .10$, and $F_2(1, 24) = 6.1$, $p < .05$, and the main effect of test position was not significant (Fs < 1.0, ps > .3). The 95% confidence interval on the response time means was 21 ms.

For probabilities correct, in Experiment 1, responses for nonreferents were significantly more accurate than responses for referents, likely due to recency, $F_1(1, 31) = 8.3$, $p < .01$, and $F_2(1, 24) = 6.1$, $p < .05$. All other probability correct effects for both experiments, including critical interactions, were not significant (Fs < 1.8, ps >.19).

The results of Experiments 1 and 2 show that an increase in accessibility occurs for the long but not the short versions of the stories. Simply increasing the lengths of the stories, it seems, made pronoun resolution successful. We attribute this to the longer stories themselves being more engaging, allowing for less superficial processing.

For Experiments 1 and 2, as for all of the probe recognition experiments presented in this paper (except Experiment 6, where responses from a single test position were collected), we report the main effects of test word and test position on our dependent variables but do not attempt to interpret them. This is because, due to our desire to use naturalistic stimuli (and to replicate Greene et al., 1992), there were a number of factors that might affect recognition response time and accuracy for which we could not control. In order to interpret a main effect of test word, for instance, we would have needed to control for the length and frequency of the character names but also for more amorphous characteristics of our stories, such as how readers' gender stereotypes may interact with the events described in the text (e.g., a story about a bakery may seem to be more "about" a female protagonist than a male protagonist, no matter how many times each is mentioned). In order to interpret a main effect of test position, we would first have to rule out uninteresting explanations, such as the placement of the test word at the beginning or the end of a line. Most important, because only the pronoun's effect on the relative accessibility of our two characters is of interest to us, we limit our discussion to interactions between test words and test positions.

Experiments 3 and 4

In Experiments 1 and 2, the test positions for the referent and nonreferent were immediately before the pronoun and at the end of the pronoun sentence. In Experiments 3 and 4, the second test position was moved closer to the pronoun, an average of only about three words after it.

Experiment 3 used the same short stories as in Experiment 1, stories for which there was no advantage for the referent over the nonreferent. However, it is possible that an earlier test position would show such an advantage. If so, it could reflect some initial, partial process of pronoun resolution that dissipated by the end-of-the-sentence test position in Experiment 1.

Experiment 4 used the same long stories as in Experiment 2. In Experiment 2, there was a speedup for the referent relative to the nonreferent from before the pronoun to the end of the sentence. There are two possible interpretations of this result. One is that resolution of the pronoun happened only at the end of the sentence, only as part of sentence "wrap-up" processes. The other possibility is that it occurred earlier, or at least began earlier, within a few words of the pronoun. Our aim in Experiment 4 was to distinguish between these possibilities.

Method

All elements of the experiments' designs, materials (including fillers), and procedures were the same as for Experiments 1 and 2, except that the test points were immediately before the pronoun or an average of 2.9 words ($SD = 0.65$) after the pronoun (and 2.6 words before the end of the sentence). There were 16 participants in Experiment 3 and 20 in Experiment 4.

Results and Discussion

Mean response times and accuracy values are presented in Table 3. The results of Experiments 3 and 4 replicated those of Experiments 1 and 2. For the short versions of the stories, Experiment 3, there was no differential speedup for the referent of the pronoun compared to the nonreferent from the test position before the pronoun to the test position after; this interaction was not signif-

Table 3
Results for Experiments 3 and 4: Mean Response Times (RTs) and Probabilities Correct

Experiment	Test word	Test position	Mean RTs (probabilities correct)
3 (short versions)	Referent	Before pronoun	763 ms (.88)
		After pronoun	786 ms (.89)
	Nonreferent	Before pronoun	731 ms (.96)
		After pronoun	763 ms (.96)
4 (long versions)	Referent	Before pronoun	879 ms (.91)
		After pronoun	867 ms (.92)
	Nonreferent	Before pronoun	866 ms (.93)
		After pronoun	905 ms (.95)

icant (Fs < 1.0, ps $> .3$). For the other effects, the main effect of test word approached but did not reach significance, $F_1(1, 15) = 3.0$, $p = .10$, $F_2(1, 24) = 3.2$, $p = .09$. The main effect of test position was significant for test items but not for participants, $F_1(1, 15) = 3.1$, $p = .09$, and $F_2(1, 24) = 4.3$, $p < .05$. The 95% confidence interval on the response time means was 20 ms.

In Experiment 2, with the long versions of the stories, response times for the nonreferent slowed from the test position before the pronoun to the end of the sentence by 55 ms. Response times for the referent sped up slightly, by 3 ms. Experiment 4 showed this same pattern: Response times for the nonreferent slowed from before the pronoun to several words after the pronoun by 39 ms, whereas response times for the referent sped up slightly, by 12 ms. This interaction was significant by participants and was nearly significant for items, $F_1(1, 19) = 5.0$, $p < .05$, and $F_2(1, 24) = 4.2$, $p = .052$. The main effects of test word and test position were not significant (Fs < 2.3, ps $> .14$). The 95% confidence interval on the response time means was 26 ms.

For Experiment 3, there was a significant effect of test word on probability correct for subjects (but not items), likely due to recency, with responses more accurate to nonreferents than referents, $F_1(1, 15) = 8.4$, $p < .05$, $F_2(1, 24) = 3.2$, $p = .09$. All other probability correct Fs, including those for the critical interactions for Experiment 3 and Experiment 4, were less than 1.0 (ps $> .3$).

These data, like those of Experiments 1 and 2, support our hypothesis that the longer stories engaged readers to a greater extent than the shorter ones. Even after only about three words, the pronoun made the referent more accessible.

Experiment 5

Experiment 5 addressed the question of what exactly about the longer versions of the stories encouraged pronoun resolution. We speculated above that the longer, more interesting versions led to less superficial processing. But the longer versions of the stories did not just increase the length and richness of the stimuli; they also increased the number of times the referent and nonreferent were mentioned. It is possible that this increase was responsible for our results. In Experiment 5, we asked whether increased length is sufficient to encourage pronoun resolution when it does not provide additional character information. Experiment 5 was identical to Experiment 2 except that the experimental stories were rewritten such that the middle four lines no longer mentioned either of the two characters (see Table 4). For the Rita story, the new lines were as follows:

> On Tuesday the magazine would go to press. It had a circulation of almost 100,000 households. The magazine mostly contained fitness-related articles. However, it occasionally published relationship advice.

The referent and nonreferent were tested in the same locations as for Experiment 2, immediately before the pronoun and at the end of the pronoun's sentence.

Method

The stories used for Experiment 5 (see Table 4) were the same as those used for Experiment 2, except that lines 3 through 6 were rewritten such that they no longer mentioned either the referent (e.g., Rita) or the nonreferent (e.g., Walter). The procedure, design, and filler stories were the same as those for Experiment 2. There were 24 participants.

Results and Discussion

The same pattern of results was found in Experiment 5 as in Experiment 2 (see Table 5): Responses for the nonreferent slowed by 58 ms from the position before the pronoun to the position at the end of the sentence, whereas responses for the referent did not (they sped up by 17 ms). Response times were slower overall in Experiment 5 than in Experiment 2. This is unsurprising, because the stories used in Experiment 5 did not reference the characters as often.

The critical interaction between test position and test word was significant, $F_1(1, 23) = 5.6$, $p < .05$, and $F_2(1, 24) = 4.7$, $p < .05$. The main effects of test position and test word were not significant

Table 4
Sample Stimulus for Experiment 5

Story	Rita and Walter were writing articles for a magazine. They had to get it done before next Tuesday. *On Tuesday the magazine would go to press.* *It had a circulation of almost 100,000 households.* *The magazine mostly contained fitness-related articles.* *However, it occasionally published relationship advice.* Rita edited the section Walter had written and then [TEST] she smoked a cigarette to relax. [TEST]
Test words	Rita (referent), Walter (nonreferent)

Note. Italics mark where stimuli differ from those used in Experiment 2.

Table 5
Results for Experiment 5: Mean Response Times (RTs) and Probabilities Correct

Test word	Test position	Mean RTs (probabilities correct)
Referent	Before pronoun	981 ms (.95)
	After pronoun	964 ms (.94)
Nonreferent	Before pronoun	959 ms (.97)
	After pronoun	1,017 ms (.99)

(Fs < 1.0, ps $> .3$). The 95% confidence interval on the response time means was 30 ms. For probability correct, there was a main effect of test word for subjects, $F_1(1, 23) = 5.2$, $p < .05$, but not for items, $F_2(1, 24) = 2.8$, $p = .11$. The main effect of test position and the interaction between test position and test word were not significant (Fs < 1.2, ps $> .28$).

These results indicate successful pronoun resolution, even though the lines added to the story did not contain any references to either the pronoun's referent or the nonreferent. This suggests that the length and richness of a story, apart from the amount of information devoted to specific characters, contribute to pronoun resolution.

Experiment 6

For Experiments 1 through 5, the pronoun's referent was always the subject, not the object, of the main clause of its sentence. For example, in the sentence "Rita edited the sections that Arthur had written and then she smoked a cigarette to relax," the pronoun *she* refers to Rita. It is often thought that subjects have a privileged status as the topic or focus of their sentences (see, e.g., Hudson, Tanenhaus, & Dell, 1986). If this is correct, the increased engagement that we hypothesized for the long versions of the stories might lead to pronoun resolution when the pronoun refers to the subject of its sentence but not when it refers to the object.

For Experiment 6, the last sentence of the long version of each story from Experiment 2 was rewritten such that the referent of the pronoun was the object of the first clause of the sentence. The "Rita" sentence, for example, became "Rita edited the section Walter had written while he smoked a cigarette to relax."

We used the long versions of the stories for this experiment because it was only with them that evidence for pronoun resolution was present. With the long versions (Experiments 2, 4, and 5), responses for the object—the nonreferent—were slower at the ends of the sentences than were responses for the referent. In Experiment 6, we looked for a reversal of this difference: Response times for the object should be faster than response times for the subject, because it is the object that is the referent of the pronoun.

Method

The long versions of the experimental materials were modified so that the referent of the pronoun in the last clause was the object of the preceding clause, as in the "Rita" sentence above. To look for the reversal, we used only the test points at the ends of the sentences. The filler items and procedure were identical to those for Experiment 2. There were 16 participants.

Results and Discussion

As predicted by the referent being correctly interpreted as the object of the first clause, response times for the object were faster than response times for the subject, by a significant 41 ms. This result shows that the advantage for the referent over the nonreferent in the earlier experiments was not dependent on the referent being the subject of its sentence.

The mean response time for the referents was 824 ms (.98 probability correct), and that for the nonreferents was 864 ms (.95 probability correct). This effect was significant, $F_1(1, 15) = 12.1$, $p < .01$, and $F_1(1, 24) = 8.3$, $p < .01$. The 95% confidence interval on the response time means was 16 ms. The difference in probabilities correct was significant with items but not with participants, $F_1(1, 15) = 2.5$, $p = .13$, and $F_2(1, 24) = 4.3$, $p < .05$.

It should be noted that the pronoun was not the only relevant change to the final line. Many of the stories had to be tweaked so that the object of the first clause could continue as the referent of the pronoun and the story would still read naturally. In our Rita and Walter example, for instance, we changed *and then* to *while*. It could be argued that *and then* may lead participants to expect the subject of the first clause to continue as the subject of the second clause, and *while* leads participants to expect the object of the first clause to become the subject of the second. Still, this does not alter our main finding, that participants show evidence of pronoun resolution (even if, as is so often the case, resolution is encouraged by other textual factors) when the stories are long and not when they are short.

Experiment 7

For the long versions of the stories, the previous experiments have demonstrated that a pronoun differentially affects the referent and the nonreferent characters, providing the referent with a relative boost. In our account, this boost occurs because the pronominal referent is properly integrated into a reader's story representation. However, we have not yet demonstrated that the pronoun is causing the referent to remain accessible longer than it otherwise would. It is also possible that the pronoun causes the nonreferent to become less accessible, perhaps because it is a direct competitor for the referent (e.g., Gernsbacher, 1989; Rigalleau et al., 2004). Experiment 7 tested between these alternatives.

Experiment 7 was identical to Experiment 2 except for the recognition test words presented to participants. Instead of comparing the referent character to the nonreferent character, we compared the referent character to a noun mentioned in the first line of the story (e.g., *article* for the Rita story). The noun was not related to either the referent or the nonreferent character other than that it occurred in the same story. Because the noun was not more related to the referent than the nonreferent, there is no reason to expect the accessibility of this test word to be influenced by the pronoun. The relative accessibility of *article* before and after the pronoun, then, can discriminate which character is affected by the presence of the pronoun. If *article* behaves like the nonreferent character, it can be concluded that the different patterns of data associated with the nonreferent and the referent characters are due

to the pronoun's effect on the referent character and are not an effect of inhibition on the nonreferent character.

Method

The only difference between Experiment 7 and Experiment 2 was that the accessibility of the referent was compared to the accessibility of a noun from the first sentence, rather than to that of the nonreferent. The referent and the noun were tested immediately before the pronoun and at the end of the pronoun sentence. The stories, procedure, and design were identical to those for Experiment 2. There were 18 participants.

Results and Discussion

The results from Experiment 7 are shown in Table 6. Responses to the noun from the first sentence of the story slowed from before the pronoun to after the pronoun, just as responses to the nonreferent character slowed in the earlier experiments (though overall accuracy for the control word was lower, likely because the control word was mentioned only once, in the first line). Response times to the noun slowed by 82 ms from the test position before the pronoun to the test position at the end of the pronoun sentence, compared with the nonreferent's 55 ms slowdown in response times in Experiment 2. Response times to the referent character, on the other hand, decreased by 24 ms. The interaction between test word (noun and referent character) and test position was significant, $F_1(1, 17) = 5.9$, $p < .05$, and $F_2(1, 24) = 4.7$, $p < .05$. We interpret this finding as evidence that the pronoun is not working to make a competitor character less accessible; rather, it is working to keep the referent character more accessible.

The main effect of test word was significant, $F_1(1, 17) = 51.5$, $p < .001$, and $F_2(1, 24) = 40.0$, $p < .001$, and the main effect of test position was marginally significant, $F_1(1, 17) = 3.1$, $p = .10$, and $F_2(1, 24) < 1.0$, $p > .30$. The 95% confidence interval on the means was 20 ms.

For probabilities correct, responses were more accurate for the referent than for the control word, $F_1(1, 17) = 45.7$, $p < .001$, and $F_2(1, 24) = 30.5$, $p < .001$. The main effect of test position and the interaction of test word and test position were not significant (Fs < 1.0, ps $> .3$).

Experiment 8

This experiment, like the earlier experiments, examined the nature of the relative increase in accessibility associated with the referent of a pronoun during pronoun resolution for the long versions of the stories. By the end of the sentence containing the pronoun, the pronoun's referent enjoys increased accessibility, but we do not know how long this boost in accessibility lasts. In Experiment 8, we tacked an additional sentence onto the end of each of the long stories used in Experiment 2. For the Rita story, it was "Tuesday was very quickly approaching." The added sentences did not contain references to either the referent or the nonreferent character. Participants were probed either immediately before the additional sentence (i.e., at the end of the sentence containing the pronoun) or immediately after the additional sentence (see Table 7).

There were three possible patterns of results. First, the accessibility of the referent character could remain high, with the accessibility of the nonreferent character continuing to decrease. Second, the accessibility of both characters could decrease, with the referent of the pronoun still enjoying some degree of relatively higher accessibility by the sentence's end. Finally, the accessibility of the referent could decay more quickly than the accessibility of the nonreferent, such that the relative accessibilities of the referent and the nonreferent characters became the same as they were before the pronoun was read.

We did not expect the first outcome, that the referent maintains a high level of accessibility while the accessibilities of other discourse entities decrease. In our account, anaphoric resolution is determined by the relative accessibility of discourse entities with respect to the anaphor. Pronouns and other forms of anaphora appear regularly in stories, and, while a referent's accessibility boost should last long enough for the referent to be connected to the rest of the discourse, it should not last long enough to interfere with other instances of anaphoric resolution. The second two outcomes seemed more plausible. For both, the relative accessibility of the referent decreases from the end of its sentence to the end of the added sentence.

Method

An additional sentence was added to the end of each story used in Experiment 2. The sentence took up a single line on the PC screen and contained no references to the referent character or the nonreferent character. An additional sentence was also added to each of the filler stories from Experiment 2.

The procedure and design were the same as for Experiment 2 except for the recognition probe test positions. The test positions were immediately after the sentence containing the pronoun (the second position in the experiments described previously) and immediately after the additional sentence. The test words remained the same as Experiment 2 (the names of the referent and the nonreferent characters). There were 32 participants.

Results and Discussion

Mean response times and probabilities correct from Experiment 8 are presented in Table 8. When the test word was the name of the character referenced by the pronoun, response times slowed from the end of the sentence containing the pronoun to the end of the added filler sentence (95 ms). The slowdown for the nonreferent was smaller, at 35 ms. This pattern is consistent with our third scenario: the accessibility of the referent decays more quickly than the accessibility of the nonreferent.

Table 6
Results for Experiment 7: Mean Response Times (RTs) and Probabilities Correct

Test word	Test position	Mean RTs (probabilities correct)
Referent	Before pronoun	851 ms (.94)
	After pronoun	827 ms (.93)
Control	Before pronoun	960 ms (.67)
	After pronoun	1,042 ms (.70)

Table 7
Sample Stimulus for Experiment 8

Story	Rita and Walter were writing articles for a magazine. They had to get it done before next Tuesday. Rita didn't trust Walter to get the facts right. Once, he'd written a piece about aliens landing in Chicago. "I'm going to get dragged down with you," Rita said at the time. However, neither of them had been fired. Rita edited the section Walter had written and then she smoked a cigarette to relax. [TEST] Tuesday was very quickly approaching. [TEST]
Test words	Rita (referent), Walter (nonreferent)

In Experiment 2, from before the pronoun to the end of the pronoun sentence, the nonreferent slowed but the referent did not. In Experiment 8, from the end of the pronoun sentence to the end of the added sentence, the nonreferent did not slow as much as the referent did (see Table 8). If we combine these two experiments, the slowdowns from before the pronoun to the end of the added sentence were roughly the same: 84 ms for the referent and 79 ms for the nonreferent. The difference is that the referent's slowdown happened only after the end of the pronoun sentence.

Statistically, the differences between the referent and nonreferent for the three test positions should make up a triple interaction: Responses for the nonreferent slowed across all three test positions, but the responses for the referent slowed only across the last two test positions. When the data from Experiments 2 and 8 were combined, this triple interaction was significant, $F_1(1, 62) = 8.7$, $p < .01$, and $F_2(1, 54) = 7.0$, $p < .05$. The 95% confidence interval for the mean response times was 32 ms.

When we analyze the data from Experiment 8 alone, the interaction between test word and test position approached but did not reach significance, $F_1(1, 31) = 3.3$, $p = .08$, and $F_2(1, 24) = 2.4$, $p = .13$. However, given the triple interaction, we conducted tests based on our a priori hypotheses: The slowdown for the referent from the end of the pronoun sentence to the end of the added sentence was significant, $F_1(1, 31) = 16.5$, $p < .001$, and $F_2(1, 28) = 9.6$, $p < .01$, but the slowdown for the nonreferent was not, $F_1(1, 31) = 2.2$, $p = .15$, and $F_2(1, 28) = 1.3$, $p = .26$. The 95% confidence interval in the mean response times was 16 ms.

For probability correct, there were no significant main effects of test word or test position and no significant interaction (Fs < 1.0, ps $> .3$).

Table 8
Results for Experiment 8: Mean Response Times (RTs) and Probabilities Correct

Test word	Test position	Mean RTs (probabilities correct)
Referent	After pronoun	871 ms (.99)
	After filler sentence	966 ms (.98)
Nonreferent	After pronoun	891 ms (.99)
	After filler sentence	926 ms (.99)

Experiments 9 and 10

The data from Experiments 1 through 8 suggest that, under appropriate circumstances, pronouns are resolved. By this we mean that they are instantiated such that readers understand that it is Rita (and not Walter, or another character) who both edited the article and smoked the cigarette. The issue addressed by Experiments 9 and 10 is whether online accessibility in a story is reflected in the memory representation for the story. In addressing this issue, Experiments 9 and 10 also provide independent evidence for our interpretation of the relationship between online accessibility and pronoun resolution.

If a pronoun has been successfully resolved, its referent should be integrated into its local textual context in the long-term memory representation of the discourse (e.g., Kintsch, 1988; Kintsch & Van Dijk, 1978; McKoon & Ratcliff, 1980b). *Rita,* for example, should be more closely linked to "smoked a cigarette" than *Walter* is.

Experiments 9 and 10 used a priming paradigm (Howard, 1985; McKoon & Ratcliff, 1980b; Ratcliff & McKoon, 1978) to examine the memory representations of the short stories from Experiment 1 and the long stories from Experiment 2. Participants read blocks of either the short stories and fillers (Experiment 9) or the long stories and fillers (Experiment 10). After each block, they were given a series of words for recognition and responded "Yes" or "No" according to whether they recognized the word from a story they had read. If the pronoun's referent was correctly incorporated into the representation of the story in memory, we would expect participants to be faster to recognize *cigarette* if it appeared immediately in the test list after the referent, *Rita,* than if it appeared after *Walter.* The results from Experiments 1 and 3 predict that the preceding test word (referent vs. nonreferent) should not affect response times for target words in the short stories (Experiment 9), but it should affect response times for target words in the long stories (Experiment 10).

Method

Materials. The experimental stories used in Experiments 9 and 10 were identical to those used in Experiments 1 and 2, respectively, except that an additional line of text was inserted before the last line of the stories (i.e., before the line containing the pronoun; see Table 9). The line for the Rita story was "about the effects of sleep deprivation on worker productivity." This line was

Table 9
Sample Stimuli for Experiments 9 and 10

Both experiments	Rita and Walter were writing articles for a magazine. They had to get it done before next Tuesday. Rita didn't trust Walter to get the facts right. Once he'd written a piece about aliens landing in Chicago.
Experiment 10	"I'm going to get dragged down with you," Rita said at the time. However, neither of them had been fired.
Both experiments	Rita edited the sections Walter had written about the effects of sleep deprivation on worker productivity and then she smoked a cigarette to relax.

added to distance the two characters from the pronoun because pilot data suggested a potential ceiling effect on the abilities of the character names to prime the target words. The target word was always a noun appearing shortly after the pronoun in the final line of the story (e.g., *cigarette*). The filler stories used in Experiment 9 were identical to the short fillers used in Experiment 1, and the filler stories used in Experiment 10 were identical to the long fillers used in Experiment 2.

Procedure and design. Participants completed one session of about forty minutes. There were 24 participants in each experiment.

Both experiments began with 16 lexical decision test items. These items were included to give participants practice with the response keys on the PC keyboard. For the actual experiments, a study-test recognition memory procedure was used. Participants were given a practice block containing four filler stories, followed by a test list that contained 30 words (15 positive, from the stories, and 15 negative, not from the stories). After this practice, participants were presented with 7 blocks of 8 stories, 4 experimental and 4 filler, each followed by a test list containing 45 words (22 positive and 23 negative). A different random order of presentation of materials was used for every second participant.

The test lists were constructed in the following manner: first, the target words (e.g., *cigarette*), one for each experimental story, were placed in positions randomly chosen between 18 and 45. Then the referent or nonreferent character name intended to prime the target word was placed in the immediately preceding test position. Finally, the remaining positive test words (all nonnames taken from the filler stories) and negative test words (including six proper names) were placed randomly in the remaining positions of the test list.

Presentation times for the stories were controlled by the experimental procedure, not the participants. This was done so that we could control the amount of time between study and test. Participants began each trial by pressing the space bar of the PC keyboard. The first line of text was presented on the screen in its entirety for an amount of time determined by multiplying the number of words in the line by 325 ms. Then it disappeared from the screen and was replaced by the next line of text, and so on. When the last line of the story disappeared from the screen, there was a 1,000-ms pause before the next story was presented. After eight stories were presented in this manner, the words *TEST LIST* appeared on the screen for 3,000 ms, and then test words were presented one at a time. Each remained on the screen until the participant made a response, pressing the */* key for "Yes" if the word had appeared in any of the studied stories and the *Z* key for "No" otherwise. Participants were instructed to respond as quickly and accurately as possible. If the response to a test word was correct, then the next word appeared after 50 ms. If the response was incorrect, than the word *ERROR* appeared on the screen for 1,500 ms before the next test word appeared.

Results and Discussion

For Experiment 9, in which participants were presented with the short stories, there was no effect of prime (referent vs. nonreferent) on the target word. Response times for the target were 769 ms when it was preceded by the name of the referent character in the test list and 767 ms when it was preceded by the name of the nonreferent character. The 95% confidence interval for the mean response times was 20 ms. The probability of a correct response was .81 in both conditions.

In contrast, for Experiment 10, with the long stories, there was a priming effect. Response times for the target were 41 ms faster when it was preceded by the name of the referent character than when it was preceded by the name of the nonreferent character (734 ms vs. 775 ms). This difference was significant, $F_1(1, 21) = 4.6$, $p < .05$, and $F_2(1, 25) = 6.4$, $p < .05$. The 95% confidence interval for the mean response times was 32 ms. The probabilities of correct responses were .84 and .88, respectively (Fs < 2.5, ps $> .12$).

The results from Experiments 9 and 10 are exactly what would be expected if online accessibility maps onto memory representation. The longer stories, which showed an online accessibility advantage for the referent character over the nonreferent character (Experiments 2, 4, 5, and 6), also showed a priming advantage for the referent character over the nonreferent character (Experiment 10). This suggests that, unlike in the shorter stories, in the longer stories the pronoun was resolved online and was also appropriately integrated into the memory representation of the story.

Experiments 9 and 10 provide corroborating evidence for our accessibility-based view of pronoun resolution. They suggest that the online accessibility of a referent affects its integration into a memory representation of the text. This in turn suggests that we are correct to use the relative accessibilities of discourse entities as a dependent variable to measure pronoun resolution.

Experiment 11

For all of Experiments 1–10, we have interpreted the results in terms of readers' engagement. Readers were more engaged with the longer stories than the shorter ones, which leads readers to set a higher standard for coherence and thus facilitates pronoun resolution. However, readers had not yet been asked whether they actually found the longer stories engaging. In Experiment 11, we presented readers with a short questionnaire designed to elicit precisely this information.

Method

Materials, procedure, and participants. Four different paper questionnaires were created, each containing 14 of the 28 total stories used in Experiments 1 and 2. Story length was manipulated within subject: Seven of the stories on each questionnaire were short versions from Experiment 1, and seven were long versions from Experiment 2. Stories of a particular length were presented in a single block. Two of the questionnaires presented the block of short stories first, and the other two presented the block of long stories first.

Participants were instructed to read carefully because they would be asked questions about the stories. On the final page of the questionnaire, participants were asked three questions intended to measure their self-reported engagement in the stories. Participants were asked (a) which set of stories (long or short) they found most interesting, (b) in which set of stories the events were easier to imagine, and (c) which set of stories better transported them into the story world. There were 24 participants in this experiment, which took about twenty minutes to complete.

Results and Discussion

For all three of the questions, participants were more likely to answer "long" than "short": Twenty out of 24 responded that the long stories were more interesting (83.3%), 16 out of 24 responded that events in the long stories were easier to imagine (66.7%), and 23 out of 24 responded that the long stories better transported them into the story world (95.8%). Pearson's chi-square tests revealed a significant effect of story length on interestingness, $\chi^2(1) = 10.7$, $p < .01$, and transportation, $\chi^2(1) = 26.2$, $p < .001$. Length approached but did not reach significance for ease of imagining, $\chi^2(1) = 2.7$, $p = .10$. We interpret the results of the questionnaire as support for our hypothesis that readers are less engaged with the shorter versions of the stories than the longer versions, leading to a more superficial understanding of who, precisely, is doing what.

General Discussion

In Experiments 1 through 4, participants responded differently to longer than shorter stories. We found an increase in accessibility for the referent character relative to the nonreferent character when a pronoun was encountered in the longer but not the shorter stories. We found this regardless of the content of the additional lines (Experiment 5); the referent and nonreferent characters did not have to appear in the added text for pronoun resolution to occur. In Experiments 1 through 5, the referent was always the subject of the pronoun's sentence. In Experiment 6, the referent was switched to the object position for the longer stories. We found that the referent test words still showed an advantage over the nonreferent test words. In Experiment 7, we determined that the difference in relative accessibility between the referent and nonreferent characters was due to a facilitory effect on the referent character, not an inhibitory effect on the nonreferent competitor.

In Experiment 8 we traced the time course of the referent character's accessibility advantage. We found evidence that the referent's accessibility, high at the end of its sentence, decreased by the end of the next sentence such that it no longer enjoyed an advantage over the nonreferent. The pronoun, then, served to boost the referent's accessibility only temporarily.

Experiments 9 and 10 provide evidence that this temporary boost allows for proper integration, actually changing how the story is remembered. In longer stories, where the referent character gets an accessibility boost after the pronoun, the referent is closely linked in memory to the pronominal discourse context; in the shorter stories, this was not the case.

Finally, in Experiment 11, we found that, when asked directly, participants agreed with us that the longer stories were more engaging: Participants rated them as more interesting, easier to imagine, and more able to transport readers into the story world.

We interpret all of these data under the assumption that levels of processing can vary. At lower levels, processing may simply ensure only that there is at least one highly accessible referent of the appropriate gender and number (Greene et al., 1992). This operates whether readers are fully engaged in a story or reading only superficially. At higher levels, pronouns are linked to referents. A referent that matches the pronoun in gender and number will become relatively more accessible after the pronoun than all other possible referents as it is integrated into the story.

A number of studies (i.e., Garnham, Oakhill, & Cruttenden, 1992; Greene et al., 1992; Rigalleau & Caplan, 2000; Rigalleau et al., 2004; Sanford & Garrod, 1989; Stewart et al., 2007) have all proposed the same sort of gender/number matching process that we do, a process that operates whether a reader does or does not have strategic goals and that does not depend on the reader's level of engagement. If there is available at least one matching entity, reading can proceed without interruption. We believe, however, that if there is no matching entity—or if there is more than one—the referent does not receive a boost in accessibility relative to the nonreferent and the connection between referent and pronoun is not made.

Indeed, as Rigalleau et al. (2004) pointed out, such assumptions may go a long way toward explaining conflicting findings regarding the automaticity of pronominal resolution across different tasks and modalities. For instance, Rigalleau and colleagues (Rigalleau & Caplan, 2000; Rigalleau et al., 2004) have used reading time studies to demonstrate that even during shallow processing conditions, readers are disrupted when pronouns are used infelicitously (e.g., when a male character is described as "she"). This suggests that pronominal gender is automatically matched against possible referents. However, detecting mismatches and fully instantiating pronouns are not one and the same. By instantiation, we mean something quite specific: that information predicated of the pronoun is connected in memory to information predicated of the referent.

For similar reasons, studies of automatic resolution across different modalities often reach seemingly incompatible conclusions. For instance, Arnold, Eisenband, Brown-Schmidt, and Trueswell auditorally presented participants with stories such as the following: "Donald is bringing some mail to Minnie while a violent storm is beginning. She's carrying an umbrella, and it looks like they're both going to need it" (2000, p. B18). Using a visual-world paradigm (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1996), Arnold et al. compared eye movements toward pictures of the two characters (Donald and Minnie) and found that shortly after the onset of the pronoun, movements toward its referent increased. The stories were not presented slowly, nor were participants asked comprehension questions that required resolution, and yet the results suggest a relatively immediate and automatic resolution process. However, because the referent and nonreferent characters both remained in the visual field throughout the story, memory demands for this task were considerably reduced. It may be the case that, with both Minnie and Donald in sight, the matching process (where features of "she" are matched to features of Minnie) may lead to instantiation (where information about the pronoun, i.e., *carries an umbrella*, is connected to information about Minnie, i.e., *recipient of mail*) more easily than when the referent must be reinstated in memory. The visual cues that would make this possible, however, are absent during reading. It may also be the case that Arnold et al.'s task was simply more engaging. This would lead to the construction of richer discourse representations (something Arnold et al. themselves suggest). Indeed, later in the General Discussion we describe another instance in which representational richness, and not necessarily length, leads to an increased probability of pronoun resolution.

We interpret our results as showing that instantiation occurs automatically when a reader's level of engagement is higher than it was with our short stories. In our experiments, the long stories added four lines to the short stories, doubling their size. The data provide two pieces of evidence for instantiation. The first comes

from the experiments with the online testing procedure. We found that the accessibility of the referent increased relative to the nonreferent from the test position before the pronoun to test positions after the pronoun. From before the *she* in "she smoked a cigarette to relax" to after it, response times to *Rita* decreased relative to those for *Walter.*

The second piece of evidence for instantiation comes from tests of long-term memory, Experiments 9 and 10. After a block of several stories, we tested single words for recognition and observed a priming effect. Responses were faster when a word predicated of the pronoun was preceded by the referent than by the nonreferent. In terms of the Rita sentence, "smoked the cigarette" is predicated of Rita, not Walter. Therefore, responses to *cigarette* were faster when it was immediately preceded in the test list by *Rita* than when it was immediately preceded by *Walter.* We interpret this in terms of Kintsch's (1988) construction–integration model. With instantiation, information about Rita is tied directly to information about the pronoun. "Rita edited the section that Walter had written" is connected directly to "Rita smoked a cigarette to relax."

A large and growing body of literature examines how language comprehenders determine which of the entities in a discourse is referenced by an ambiguous pronoun. Possible referents in the subject position of a clause (e.g., *Samantha* in "Samantha told Jenna all about the family she grew up with") and possible referents in a parallel position with respect to the anaphor (e.g., *Steven* in "James helped Steven with his English paper and then Dan helped him with his math exercises") have been shown to have an advantage over other discourse entities (Crawley et al., 1990; Smyth, 1994). Furthermore, pragmatic clues have been shown to guide a comprehender's focus to one possible referent (Arnold, 2001; Marslen-Wilson et al., 1993; Rohde et al., 2006; Stevenson et al., 1994). For instance, in a sentence such as "Janice bought Helen a card because she had a birthday coming up," real-world knowledge of causal relationships and birthdays supports the conclusion that it is Helen, not Janice, with a nearby birthday.

As compelling as such research is, it does not take into account the possibility that, in some situations, not only ambiguous pronouns but also unambiguous pronouns are not resolved. As mentioned in the introduction above, psycholinguistic research offers a growing body of support that readers can engage in superficial levels of processing (e.g., Christianson et al., 2001; Ferreira et al., 2002; McKoon & Ratcliff, 1992). In many ways, this is intuitive: We do not read a magazine advertisement with the care and precision with which we would read an anatomy textbook or the enthusiasm with which we would read the last chapter of a *Harry Potter* novel. Thus, as our experiments elucidate, connections made so quickly and easily as to be considered automatic under some conditions may not be made at all under others.

We interpreted the results of Experiments 1 and 2 in terms of readers' engagement, with longer, more interesting stories encouraging increased engagement with the texts. Busselle and Bilandzic (2008) recently developed a model of story comprehension and engagement. They argued that there exists a strong relationship between the cohesiveness or consistency of a narrative, both internal (with itself) and external (with the world), and reader engagement. We argue that a longer text has more potential for internal and external consistency than does a shorter one, simply because the text has more room with which to establish and adhere to a set of rules or a context. For instance, if an author has established that some but not all of the laws of physics apply in the story world or that a certain character is a daredevil, a longer text gives the author more room to provide details in support of these claims.

Bilandzic and Busselle (2008) have further argued, "To the extent that constructing a storyworld occupies cognitive resources, the audience member must give up consciousness of his or her actual surroundings" (p. 263). Readers are more likely to be transported by a story to which they must allocate more attention, which a longer, more complex story would presumably require. Assuming a limited pool of cognitive resources, Busselle and Bilandzic predicted that increased narrative engagement would lead to impaired performance in a secondary task. They reported a study in which participants who were engaged in a primary task, watching television, were also asked to press a button in response to an auditory cue. Participants' cued response times were slower in more suspenseful scenes than in less suspenseful scenes (though the emotional intensity of the scenes did not affect response times). In our experiment, interestingly, response times to test words were slower overall for the longer than the shorter stories (878 ms vs. 836 ms when collapsing across all four conditions). We point this out because it provides tentative support for our engagement hypothesis. However, we caution that the increased probability of making a correct "Yes" response to test words while reading the shorter stories could also be responsible for this difference.

Another possible explanation for our results, not yet discussed, is that, just as it takes a few moments to get into the groove of jogging or the rhythm of dancing, it may take a few moments to start up some of the cognitive processes involved in natural, fluid reading. It may be the case that many seemingly automatic processes are in fact only automatic after participants have constructed a sufficiently extensive representation of a text. The precise amount of text this requires would probably vary by text and by reader, but it seems possible that, until some sufficient representation is established, readers do not properly integrate new information.

However, new research by Johns, Long, and Swaab (2010) suggests that this "start-up" hypothesis is not correct. In an electroencephalographic study, Johns et al. found that simply providing participants with higher quality referents and nonreferents (e.g., Bill Clinton and John Travolta instead of Bill Smith and John Jones) facilitated processing for coherent sentences and further disrupted processing for incoherent sentences (e.g., those referencing *Bill Clinton* with *she*). This suggests that the richness of the discourse and not its extensiveness per se may have been responsible for improved pronoun resolution in our longer stories (see also Arnold et al., 2000).

Our experiments show that there is a qualitative difference between what readers experience and take away from shorter and longer stories: When readers are presented with short stories, they may engage in incomplete processing. And the short stories used in our experiments are by no means unique. The tasks and stimuli used in some traditional psycholinguistic experiments have been shown by a number of researchers to encourage shallow processing (Ferreira et al., 2002; Sanford, 2002; Sanford & Sturt, 2002). We should take care, when creating and interpreting experiments, to remember that reading is not necessarily comprehending.

Another question that has not yet been addressed is what, if anything, is being integrated into a discourse representation when a pronoun is not resolved. Presumably, readers are integrating something into the discourse representation. Klin et al. (2006) found evidence that referents of unresolved nominal anaphors are, nonetheless, partially encoded. Reading times were faster for sentences that contained a nominal anaphor (e.g., dessert) when the earlier discourse included a possible referent (e.g., tart), even though readers were no faster at responding to *dessert* in a lexical decision task when it was preceded by *tart* (i.e., there was no priming). Klin et al. interpreted this as evidence that readers treat an unresolved nominal anaphor as having some but not all features of the referent.

Although pronominal referents contain much less information than nominal referents, readers may partially encode them as well. Readers may integrate the information a pronoun does contain into the "slot" in the textual representation that the pronoun's referent might usually fill. Take, for instance, this sentence from our experiments: "Rita edited the section that Walter had written and then she smoked a cigarette to relax." It seems reasonable that, if Rita is not specifically linked to the discourse, an incomplete representation is linked instead, one that may include the grammatical properties of the pronoun itself (e.g., gender and number). Another possibility is that readers do not link anything at all into the referent's slot in the textual representation. Readers may keep this slot empty, perhaps to be filled by further information. A third possibility is that readers do integrate the pronoun's referent into a textual representation, but that it is integrated too weakly for evidence of resolution to be obtained. It is our hope that future research will distinguish among these possibilities.

The research described in this article provides both online and offline support for our account that anaphoric resolution is determined by the relative accessibilities of discourse entities with respect to an anaphor. The anaphoric referent, if determined, benefits from a temporary boost in accessibility. Furthermore, we claim that short stories are read differently than longer stories and, specifically, that when a story is too short (or sufficiently lacking in richness) to permit full engagement, readers may not fully resolve the unambiguous pronouns that, in longer texts, appear to be resolved automatically.

References

Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes, 31,* 137–162. doi:10.1207/S15326950DP3102_02

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76,* B13- B26. doi:10.1016/S0010-0277(00)00073-1

Bilandzic, H., & Busselle, R. (2008, May). *Attention and narrative engagement: Divergences in secondary task reaction times and self-reports of narrative engagement.* Paper presented at the Convention of the International Communication Association, Montreal, Ontario, Canada.

Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory, 18,* 255–280. doi:10.1111/j.1468-2885.2008.00322.x

Carver, R. (1989). *Where I'm calling from: Selected stories.* New York, NY: Vintage Books.

Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42,* 368–407. doi:10.1006/cogp.2001.0752

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, 19,* 245–264. doi:10.1007/BF01077259

Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 121–132. doi:10.1016/S0022-5371(83)80010-3

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20,* 540–551. doi:10.1016/S0022-5371(81)90165-1

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11–15. doi:10.1111/1467-8721.00158

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. doi:10.1080/01690969208409386

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition, 32,* 99–156. doi:10.1016/0010-0277(89)90001-2

Gerrig, R., & O'Brien, E. (2005). The scope of memory-based processing. *Discourse Processes, 39,* 225–242. doi:10.1207/s15326950dp3902&3_7

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67. doi:10.1037/0033-295X.91.1.1

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79,* 701–721. doi:10.1037/0022-3514.79.5.701

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 266–283. doi:10.1037/0278-7393.18.2.266

Grosz, B. J. (1981). Focusing and description in natural language dialogues. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 84–105). Cambridge, England: Cambridge University Press.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. *Proceedings of the 21st Annual Meeting of the Association of Computational Linguistics* (pp. 44–50). Cambridge, MA: Association of Computational Linguistics.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of the discourse. *Computational Linguistics, 12,* 175–204.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16,* 96–101.

Howard, D. V. (1985, August). *Aging and episodic priming: The propositional structure of sentences.* Paper presented at the meeting of the American Psychological Association, Los Angeles, CA.

Hudson, S. B., Tanenhaus, M. K., & Dell, G. S. (1986). The effect of the discourse center on the local coherence of a discourse. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 96–101). Hillsdale, NJ: Erlbaum.

Johns, C. L., Long, D. L., & Swaab, T. Y. (2010, March). *Do you know who that is? Real-world reference and coreferential processing.* Poster session presented at the City University of New York Conference on Human Sentence Processing, New York, NY.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Newton, MA: Allyn & Bacon.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction–integration model. *Psychological Review, 95,* 163–182. doi:10.1037/0033-295X.95.2.163

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394. doi:10.1037/0033-295X.85.5.363

Klin, C. M., Guzmán, A. E., Weingartner, K. M., & Ralano, A. S. (2006). When anaphor resolution fails: Partial encoding of anaphoric inferences. *Journal of Memory and Language, 54,* 131–143. doi:10.1016/j.jml.2005.09.001

Levine, W. H., Guzmán, A. E., & Klin, C. M. (2000). When anaphor resolution fails. *Journal of Memory and Language, 43,* 594–617. doi:10.1006/jmla.2000.2719

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Marslen-Wilson, W. D., Tyler, L. K., & Koster, C. (1993). Integrative processes in utterance resolution. *Journal of Memory and Language, 32,* 647–666.

McKoon, G., & Ratcliff, R. (1980a). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668–682. doi:10.1016/S0022-5371(79)90255-X

McKoon, G., & Ratcliff, R. (1980b). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior, 19,* 369–386. doi:10.1016/S0022-5371(80)90267-4

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 108–115. doi:10.1037/0278-7393.12.1.108

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99,* 440–466. doi:10.1037/0033-295X.99.3.440

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626. doi:10.1037/0033-295X.89.6.609

Myers, J. L., & O'Brien, E. J. L. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26,* 131–157. doi:10.1080/01638539809545042

O'Connor, F. (2001). A good man is hard to find. In J. H. Pickering (Ed.), *Fiction 100: An anthology of short stories* (pp. 1135–1146). Upper Saddle River, NJ: Prentice Hall.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17,* 403–417. doi:10.1016/S0022-5371(78)90238-4

Rigalleau, F., & Caplan, D. (2000). Effects of gender marking in pronominal coindexation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53*(A), 23–52. doi:10.1080/027249800390655

Rigalleau, F., Caplan, D., & Baudiffier, V. (2004). New arguments in favour of an automatic gender pronominal process. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 57*(A), 893–933.

Rohde, H., Kehler, A., & Elman, J. (2006). Event structure and discourse coherence biases in pronoun interpretation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 697–702). Red Hook, NY: Curran Associates.

Sanford, A. J. (2002). Context, attention, and depth of processing during interpretation. *Mind & Language, 17,* 188–206. doi:10.1111/1468-0017.00195

Sanford, A. J., & Garrod, S. C. (1989). What, when and how? Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes, 4,* 235–262. doi:10.1080/01690968908406369

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Science, 6,* 382–386. doi:10.1016/S1364-6613(02)01958-7

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26,* 299–323. doi:10.1207/s15326985ep2603&4_5

Schiefele, U. (1996). Topic interest, text representation, and quality of experience. *Contemporary Educational Psychology, 21,* 3–18. doi:10.1006/ceps.1996.0002

Schiefele, U., & Schreyer, I. (1994). Intrinsische Lernmotivation und Lernen. Ein berblick zu Ergebnissen der Forschung [Intrinsic motivation of learning and learning: An overview of results of research]. *Zeitschrift für Pädagogische Psychologie, 8,* 1–13.

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior, 27,* 1–17.

Sidner, C. (1983a). Focusing and discourse. *Discourse Processes, 6,* 107–130. doi:10.1080/01638538309544558

Sidner, C. (1983b). Focusing in the comprehension of definite anaphora. In M. Brady & R. Berwick (Eds.), *Computational models of discourse* (pp. 267–330). Cambridge, MA: MIT Press.

Smyth, R. H. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research, 23,* 197–229. doi:10.1007/BF02139085

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus, and the representation of events. *Language and Cognitive Processes, 9,* 519–548. doi:10.1080/01690969408402130

Stewart, A. J., Holler, J., & Kidd, E. (2007). Shallow processing of ambiguous pronouns: Evidence for delay. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 60,* 1680–1696. doi:10.1080/17470210601160807

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

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–373). Hillsdale, NJ: Erlbaum.

Webber, B. (1983). So what can we talk about now? In M. Brady & R. Berwick (Eds.), *Computational models of discourse* (pp. 331–371). Cambridge, MA: MIT Press.

Received June 1, 2010
Revision received December 7, 2010
Accepted December 22, 2010 ■