Underspecification of syntactic ambiguities:
Evidence from self-paced reading

BENJAMIN SWETS
SUNY Stony Brook, Stony Brook, New York

TIMOTHY DESMET
Ghent University, Ghent, Belgium

CHARLES CLIFTON, JR.
University of Massachusetts, Amherst, Massachusetts

AND

FERNANDA FERREIRA
University of Edinburgh, Edinburgh, Scotland

Syntactically ambiguous sentences are sometimes read faster than disambiguated strings. Models of parsing have explained this tendency by appealing either to a race in the construction of alternative structures or to reanalysis. However, it is also possible that readers of ambiguous sentences save time by strategically underspecifying interpretations of ambiguous attachments. In a self-paced reading study, participants viewed sentences with relative clauses that could attach to one of two sites. Type of question was also manipulated between participants in order to test whether goals can influence reading/parsing strategies. The experiment revealed an ambiguity advantage in reading times, but only when participants expected superficial comprehension questions. When participants expected queries about relative clause interpretation, disambiguating regions were inspected with more care, and the ambiguity advantage was attenuated. However, even when participants expected relative clause queries, question-answering times suggested underspecified representations of ambiguous relative clause attachments. The results support the construal and "good-enough" models of parsing.

Many models of language comprehension assume that globally ambiguous sentences are more difficult to process than unambiguous sentences. For instance, some constraint-based parsing models (MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey & Tanenhaus, 1998; Tabor & Tanenhaus, 1999) posit that sentences with syntactic ambiguities generally take longer to read than their unambiguous counterparts (see Spivey-Knowlton & Sedivy, 1995, for a particularly clear statement of this claim). According to these models, multiple structures are incrementally activated during sentence comprehension. These interpretations compete to become the final one. When the sentential and contextual information being considered fails to produce a clear winner, such as with ambiguous strings, the various interpretations engage in a disruptive competition in which constraints from multiple levels of representation must be considered and weighted over a longer period of time before the system settles on a chosen interpretation. In sum, ambiguous strings in this view typically slow down reading.[1]

However, it has been difficult to demonstrate processing costs for syntactic ambiguity itself, rather than for difficulties caused when early information favors one analysis and later information favors a different analysis (Clifton, Staub, & Rayner, 2007; Gibson & Pearlmutter, 2000; Lewis, 2000). Furthermore, this ambiguity disadvantage assumption has been challenged by a line of studies showing that some syntactic ambiguities can actually lead to shorter processing times (Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001). In each of these studies, participants read ambiguous sentences and disambiguated variants while their eye movements were monitored. The readers consistently displayed ambiguity advantages in these studies. For example, Traxler et al. studied a form of relative clause ambiguity (see Sentences 1–3 below) in which a relative clause modifies either a head noun (*maid/son*) or an adjunct of the head noun (*princess*). Traxler et al. showed that it took less time for people to read ambiguous sentences, such as Sentence 1,

B. Swets, bswets@notes.cc.sunysb.edu

than to read sentences with temporary ambiguities that were later disambiguated (Sentences 2 and 3).

(1) The maid of the princess who scratched herself in public was terribly humiliated.

(2) The son of the princess who scratched himself in public was terribly humiliated.

(3) The son of the princess who scratched herself in public was terribly humiliated.

Van Gompel et al. (2001) used a similar paradigm, except that the difference between ambiguous sentences (4) and unambiguous sentences (5) centered on whether the prepositional phrase (*with the rifle*) could either modify the sentential object (*the poacher*) or be interpreted as an instrument of the verb (*killed*).

(4) The hunter killed the dangerous poacher with the rifle not long after sunset.

(5) The hunter killed the dangerous leopard with the rifle not long after sunset.

This study (see also van Gompel et al., 2005) yielded the same result as Traxler et al. (1998): Participants spent more time reading sentences disambiguated by discourse context or a syntactic cue than they spent reading globally ambiguous sentences.

One way to explain this advantage for ambiguous sentences is to assume a probabilistic race model of sentence comprehension (van Gompel et al., 2001). Race models maintain that sentences with unresolved syntactic ambiguities are read quickly because the analysis that is computed initially is never overturned at a later reanalysis stage; the first choice can always be used, since no information disconfirms it. This is not the case, however, when the ambiguity is resolved at some later point in the sentence. Consider the relative clause ambiguity examples above (Sentences 1–3). Given the somewhat equibiased nature of this ambiguity (see Swets, Desmet, Hambrick, & Ferreira, 2007), we can assume that some readers will initially take the head noun (*the maid/the son*) as the attachment site for the relative clause, and about as many other readers will select the adjunct of the head noun (*the princess*) as the attachment site. When a reader arrives at the reflexive pronoun *himself* in Sentence 2 or *herself* in Sentence 3, the gender marking of the reflexive pronoun makes clear that the relative clause string beginning with *who scratched* modifies *the son* (N1, or high attachment) and *the princess* (N2, or low attachment), respectively. This means that high (N1) attachers will be forced to reanalyze in Sentence 3, and low (N2) attachers will be forced to reanalyze in Sentence 2. Averaged across all readers, the result is an overall increase in reading times in these two conditions. However, in the ambiguous condition (Sentence 1), the reflexive pronoun *herself* is consistent with both attachment sites. This means that neither high nor low attachers will be forced to reanalyze. Independent of whether readers initially followed the N1 or the N2 interpretation, the rest of the sentence will always be consistent with their reading. Hence, reading times for ambiguous sentences will be faster than for disambiguated sentences.

Although the probabilistic race model provides a compelling account of the ambiguity advantage, other explanations can also be entertained. One is that readers simply do not commit to a particular meaning when they are not provided with clearly disambiguating information. We will refer to this as an *underspecification* account. Previous studies have offered little information about the final interpretations that readers are left with once a sentence has been read. Participants in Traxler et al. (1998) and van Gompel et al. (2001; van Gompel et al., 2005) were questioned about their final interpretations of the sentences only on occasional trials, and the questions did not assess whether or how the participants resolved the attachment of the relative clause. In addition, data about question-answering time and accuracy were not consistently collected. Crucially, race models, as well as some constraint-based competition models, assume that globally ambiguous sentences are assigned a definitive interpretation, just as disambiguated sentences are. In contrast, underspecification models assume that the final interpretation of a globally ambiguous sentence (one disambiguated neither by discourse context nor by syntactic cues) is agnostic between the meanings corresponding to the two attachments of the relative clause. In other words, using our same example, the underspecification approach assumes that the comprehender will understand the sentence to mean that someone scratched herself in public, but the comprehender will leave undecided whether that person was the maid or the princess. Because this underspecification account allows the comprehension system to sometimes compute interpretations that are shallow and incomplete, we view it as consistent with what has been termed the "good-enough" approach to language comprehension (Ferreira, Bailey, & Ferraro, 2002; Sanford & Sturt, 2002).

Syntactic Underspecification

Recent studies (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, 2003) in which comprehenders were asked directly about their understanding of sentences with a single specified meaning have demonstrated that often interpretations are based either on incomplete reanalysis (Christianson et al., 2001) or on shallow processing (Ferreira, 2003). In a self-paced reading study (Christianson et al., 2001), after participants read sentences such as *While Anna dressed the baby played in the crib*, they were asked two kinds of questions. The first kind tested whether they understood that the baby played in the crib. People answered this kind of question correctly. The second kind of question tested whether the incorrect interpretation based on the garden path (that Anna dressed the baby) remained, even after this interpretation had been disallowed by thematic-role-assigning principles. For this type of question, participants were surprisingly inaccurate. These results indicate that even though syntactic reanalysis may be successful, the semantic consequences of the original misanalysis may tend to linger. The result is a final interpretation inconsistent with the sentence's actual global structure.

Implausible passive sentences are also often misunderstood (Ferreira, 2003). Participants asked to identify

agents or patients of active, transitive sentences such as *The dog bit the man* or *The man bit the dog* performed equally well in either case. However, when the sentence was in the noncanonical, passive form, performance for implausible sentences was poor. When implausible information was presented in a passive sentence (*The dog was bitten by the man*), the syntactic analysis of the sentence competed with the meaning available from world knowledge. Thus, the products of algorithmic syntactic processing may be outweighed by information obtained from world knowledge.

The results of these studies, among many others (e.g., Glenberg, Wilkinson, & Epstein, 1982; Otero & Kintsch, 1992; Reder & Kusbit, 1991; see also Volume 42, Number 2, of *Discourse Processes* for additional articles developing this topic), illustrate that people often take short-cuts in language processing. This work has given rise to approaches to language that acknowledge the occasional incompleteness of language processing. These approaches include the "good-enough" approach to language (Ferreira et al., 2002) and the view that language comprehension is, at times, shallow (Sanford, 2002; Sanford & Sturt, 2002). A focus of these approaches is that sentence comprehension draws on a limited pool of resources, including working memory, attention, and time. Hence, people often prefer to leave ambiguities unresolved when resources are in short supply or when motivation to undertake all the necessary processing steps is low.

Good-enough theories view the results of Traxler et al. (1998) and van Gompel et al. (2001; van Gompel et al., 2005) as showing that readers shallowly interpret relative clause attachments when disambiguation is not provided by context, content, or grammar. The assumption is that it takes time to choose an attachment site and insert an ambiguous phrase in the correct location in the phrase marker, so not having to do so should result in a reading time advantage for globally ambiguous sentences. However, van Gompel et al. (2001) rejected this "laziness" account, suggesting that it is less parsimonious than a probabilistic race account (see Traxler et al., 1998, for additional discussion of such an account). van Gompel et al. (2001) argued that the laziness account requires the assumption that ambiguous and unambiguous sentences are processed differently. A reader cannot know that the incorrect analysis of an unambiguous attachment is anomalous without semantically interpreting the attachment, which entails that all sentences are semantically interpreted. However, it is possible that a processor operating in a "good-enough" manner could semantically interpret an adjunct attachment deeply enough to reject it (at some processing cost) if it is anomalous, without necessarily processing attachments deeply enough to make a choice when both are semantically acceptable. The construal model (Frazier & Clifton, 1996) makes a similar prediction. According to construal, ambiguous adjuncts are not immediately attached to a particular head, but are indeterminately associated with the thematic domain of the entire complex noun phrase. Various sources of information are then consulted to determine an interpretation. When disambiguation is not required by task demands, this consultation could be superficial: just enough to identify inappropriate interpretations, but not enough to fully disambiguate between two appropriate interpretations. The present study is designed to investigate the possibility that the language processor does not make fully specified attachment decisions in the absence of syntactic information specifying how a phrase should be attached, and, more generally, to provide more information about the processing of globally ambiguous sentences.

So far, we have discussed how the lack of adequate sentence comprehension questions in previous research (Traxler et al., 1998; van Gompel et al., 2001) raises the possibility that readers failed to decide on a single attachment for ambiguous sentences. More broadly, this concern assumes that the task the participant is asked to perform after reading sentences can influence how deeply those sentences are processed. The tendency to try to resolve the attachment definitively depends on whether people are asked questions about the attachment. Readers might be less apt to fully resolve an ambiguity if they are never queried about their interpretation of the ambiguity, and more likely to dedicate resources to resolution if they believe they will be accountable for their interpretation of the sentence. The present experiment was intended to explore this broader theoretical issue of goal-related and strategic language processing as well.

The Present Experiment

To address these questions, we replicated the relative clause attachment study of Traxler et al. (1998) using a self-paced reading task. In addition, we tried to influence the depth of processing by manipulating the number and difficulty of the comprehension questions that followed the sentences. In previous studies of the ambiguity advantage (Traxler et al., 1998; van Gompel et al., 2005; van Gompel et al., 2001), readers were sometimes asked questions about final sentence interpretation (half of the experimental trials in Traxler et al., 1998, and van Gompel et al., 2005; one third of the experimental trials in van Gompel et al., 2001). Accuracy rates were only sometimes reported in the various studies (only in van Gompel et al., 2005; van Gompel et al., 2001). Response times, on the other hand, were not reported at all, as is typical in studies of syntactic processing (including our own previous work). Most importantly, the questions used in earlier studies did not ask for interpretations of relative clause attachment. Usually, the questions focused on aspects of meaning related to the matrix clause.

In the present experiment, we had three question-answering conditions. One third of the participants were assigned to the *relative clause* condition. They answered questions on every trial, and those questions were always about the interpretation of the relative clause. They also answered mildly challenging specific questions about filler items. Another third of the participants were assigned to the *superficial* condition. On every trial, they answered superficial questions that minimally tested sentence comprehension and that were never about the relative clause attachment. The final third of the participants occasionally received questions, and these were always

superficial (the *occasional superficial* condition). The latter condition is the most similar to what has been done in previous studies on this question (Traxler et al., 1998; van Gompel et al., 2005; van Gompel et al., 2001).

Two types of hypotheses are being tested. According to obligatory-attachment models, which include race models (van Gompel et al., 2005; van Gompel et al., 2001; Weber-McRoy & Hirst, 1990) and the original garden path model (Frazier, 1979), syntactic attachment decisions are automatically and incrementally computed online. Although the obligatory-attachment hypothesis could certainly be formulated in a way that predicts that overall reading speed should be influenced by the frequency and difficulty of the questions, nothing in its various forms predicts that task demands will affect the pattern of reading times between the ambiguous and disambiguated conditions. For instance, the race model would continue to predict the fastest reading times for ambiguous sentences, even if overall reading rate varied with task demands. Another prediction of these accounts is based on their assumption that readers reach a final interpretation for both ambiguous and unambiguous sentences, even with minimal task demands. Following a logic extensively used in studies of discourse processing (Singer, 2006), to the effect that fast question-answering responses indicate that the questioned material was encoded or inferred during initial reading, the race model predicts that questions should be answered quickly following both ambiguous and disambiguated items (see the General Discussion for more on this issue).

The obligatory-attachment hypothesis contrasts with various formulations of a "principled underspecification" hypothesis, such as in the good-enough theory (Ferreira et al., 2002; Sanford & Sturt, 2002) and the construal theory (Frazier & Clifton, 1996). Principled underspecification accounts hold that parsers may leave ambiguous adjunct attachments underspecified. Since underspecification can be graded (that is, it may happen in some comprehension circumstances and not in others), these accounts permit task demands (e.g., question type manipulations) to influence the commitment to a definitive attachment, and consequently to affect reading time patterns for ambiguous and unambiguous sentences. Participants who receive difficult relative clause questions are accountable for their attachment of the relative clause, and might therefore be more inclined to attach it definitively into the overall structure. This tendency could eliminate the advantage for globally ambiguous sentences. In addition, to the extent that readers do not initially resolve the relative clause attachment ambiguity even when they are frequently questioned about it, principled underspecification hypotheses predict that people who have just read a sentence with an ambiguous attachment should be slower to answer a relative clause comprehension question about it than people who have just read a disambiguated sentence. This prediction follows because ambiguous sentences do not receive an interpretation online; therefore, building a propositional analysis that would allow the question to be answered must be done offline, during the question-answering phase.

METHOD

Participants

The participants were 144 students at Michigan State University who were native speakers of English. In exchange for their participation, some participants received partial credit in their introductory cognitive psychology courses, and the rest were paid $7.

Materials

Design and Stimuli. The experiment employed a 3 (sentence type) $\times$ 3 (question type) mixed design. Sentence type (see Sentences 1–3, repeated here) was manipulated within participants: Relative clauses were either fully ambiguous (1), disambiguated to attach to N1 (2), or disambiguated to attach to N2 (3).

(1) The maid of the princess who scratched herself in public was terribly humiliated.

(2) The son of the princess who scratched himself in public was terribly humiliated.

(3) The son of the princess who scratched herself in public was terribly humiliated.

We constructed 36 triplets of items on the basis of those used in Experiment 2 of Traxler et al. (1998), with some words adapted for American readers (e.g., *vicar* was changed to *priest*). All experimental items consisted of a complex noun phrase (e.g., *The maid/son of the princess*), a relative clause with a disambiguating reflexive pronoun (e.g., *who scratched himself/herself in public*), and a matrix verb phrase (e.g., *was terribly humiliated*). The sentence region where disambiguation might occur was the *reflexive pronoun*, which disambiguated the attachment of the relative clause in favor of N1 or N2, or did not disambiguate (ambiguous). We also included 106 filler items.

Question type was manipulated between participants ($n = 48$) for each condition: relative clause questions/all trials, superficial questions/all trials, and superficial questions/occasional trials. Two types of yes/no questions were created for each of these items and the filler items. One type of question probed participants' interpretation of the relative clause in the experimental sentences (e.g., *Did the maid/princess/son scratch in public?*). The other type of question probed participants' understanding of the sentences at a very superficial level (e.g., *Was anyone humiliated/proud?*). In a third condition, participants answered the superficial question type, but only on 1 randomly chosen trial out of 12. Half of the questions of each type were correctly answered with a "yes" response, and half with a "no" response. Affirmative and negative answers were counterbalanced across the lists to control for the type of answer. The experimental items and the associated questions are listed in the Appendix.[2]

Filler items were identical for each of the lists. However, the questions asked after filler sentences in the superficial question condition were written to be more general than those in the relative clause question condition. A close read was required to answer the filler questions in the relative clause condition, but only a basic grasp of the meaning of the filler item was needed to answer questions in the superficial condition. Only 1 item in 10 was followed by a superficial question in the occasional superficial question condition.

On the basis of this 3 $\times$ 3 design, nine lists of sentence/question items were prepared. Each participant viewed only one of the lists. The condition for each experimental item rotated across lists, so that each experimental item appeared once in each cell of the within-participants design. Six items from each cell of this design appeared in a list. Experimental and filler items were presented in a unique pseudorandom order for each participant, constrained so that two experimental items did not occur in sequence.

We also created three sets of practice sentences. Each set used the same 10 sentences, but had different follow-up comprehension questions to mimic the different conditions of the between-participants manipulation of question conditions.

Apparatus. The experiment was carried out on a Dell computer using E-Prime software.

Procedure

Participants were seated in front of a computer screen. Each participant received a consent form to fill out and a set of written instructions. A practice session followed in which the participants became familiar with the task by reading 10 practice sentences. Each participant pressed a mouse button to start reading sentences, which were presented in a self-paced, noncumulative, word-by-word fashion, beginning from the left edge of the screen. The first screen that participants saw outlined a sentence using a series of underscores as a sentence outline. Underlines were replaced by words one at a time, then reverted back to underlines, as participants tapped either mouse button and read along. Sentences too long to be presented on one line were presented on two lines, with the second line also beginning on the left edge of the screen. When the sentence outline disappeared, it was followed by a question that was centered in the screen. Participants answered the question by pressing the left mouse button for "yes" or the right button for "no." Response times and question-answering accuracy were recorded. When the question disappeared, the next sentence was presented. The experiment lasted 30 min.

RESULTS AND DISCUSSION

Figure 1 shows region-by-region reading times, in milliseconds per word, in the three question type conditions, beginning with the first word (*The*) and ending with the matrix verb of each experimental item. It is apparent that reading times per word were fastest toward the beginnings of sentences in all question conditions, and that readers began to slow down as they encountered the relative clause. Readers in the relative clause question condition were especially slow, beginning with the relative pronoun (*who*) and continuing through the matrix verb. (The subordinate verb region—*Verb* in Figure 1—and the prepositional phrase region—*PP* in Figure 1—are pooled across one or more words, the number of which varied across items. Reading times for those regions are shown in the figure in milliseconds per word, but statistical analyses were performed using raw total reading times per region.)

However, effects of sentence type on reading times did not become evident until the disambiguating region [$Fs < 1$ for all regions previous to the disambiguation region, except for the *who* region: $F_1(2,282) = 1.64$, $MS_e = 6,948$, $p = .20$; $F_2(2,70) = 1.67$, $MS_e = 2,020$, $p = .20$]. In the immediately preceding subordinate verb region, there was an effect of question type by participants [$F_1(2,141) = 4.60$, $MS_e = 144,204$, $p < .05$] but not by items [$F_2(2,105) = 1.21$, $MS_e = 438,184$, $p = .30$]. However, there were no effects of sentence type ($Fs < 1$).

Because there was no evidence for effects of the ambiguity manipulation in other regions, and because fully analyzing other regions does not test any hypothesis-relevant predictions, reading times were fully analyzed only in two critical sentence regions. The first of these was the disambiguating *reflexive* pronoun (which did not disambiguate in the ambiguous condition). The second was the prepositional phrase that followed the reflexive: the *postdisambiguation* region, for which all words were cumulated into a single region for analysis. This was done because in self-paced reading studies, it is often observed that reading time effects spill over into the subsequent region (see, e.g., Ferreira & Henderson, 1990).[3]

Disambiguation Region

Figure 2 displays reading times in the disambiguation region as a function of question type and sentence type. Two effects can be seen. First, readers took longer in this region if they were asked relative clause attachment questions on every trial. Second, readers in all question type conditions spent more time in this region when the pronoun disambiguated in favor of N1 than when it disambiguated in favor of N2 or did not disambiguate.

A 3 (sentence type) × 3 (question type) mixed ANOVA was conducted on the reading times for the reflexive pronoun to verify these observations. First, the factors did
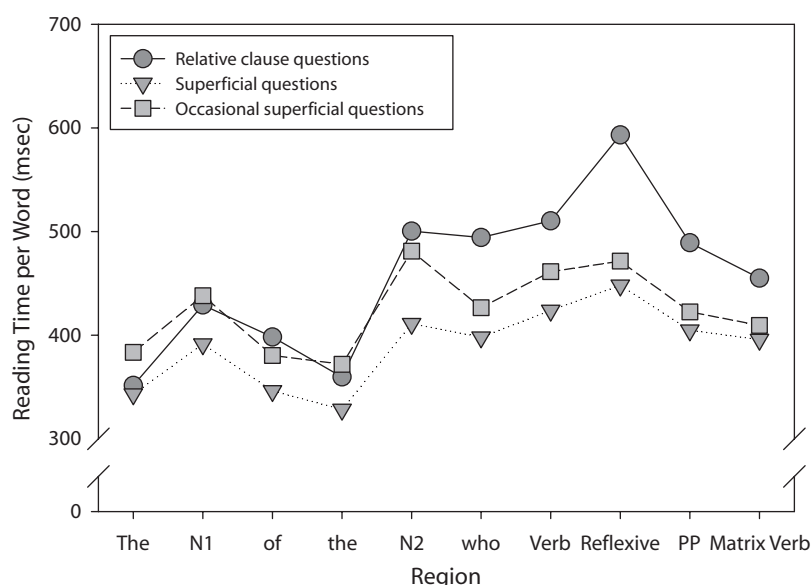

Figure 1. Reading times as a function of question type. Times are presented region by region, in milliseconds per word.
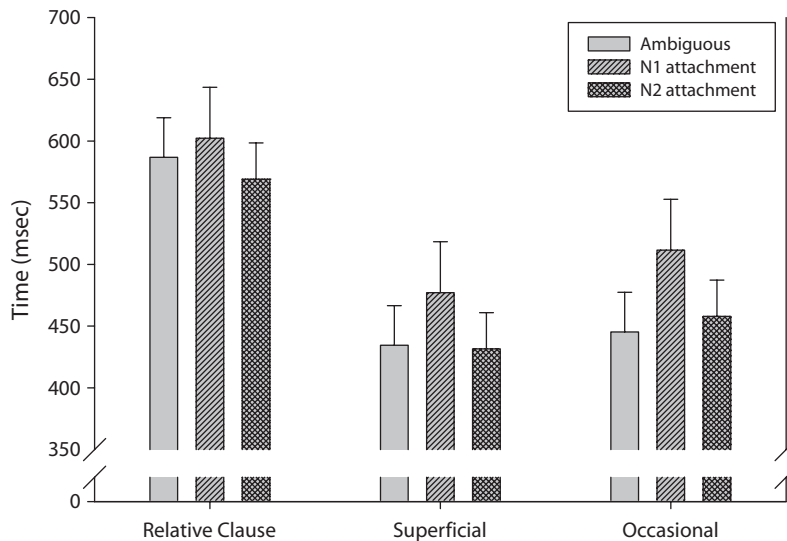
Figure 2. Reading times for the disambiguating region (e.g., *herself*) as a function of question type and sentence type. Error bars represent standard errors.

not interact ($Fs < 1$). The ANOVA revealed significant main effects of the between-participants variable of question type [$F_1(2,141) = 5.56$, $MS_e = 283,261$, $p < .01$; $F_2(2,105) = 24.94$, $MS_e = 57,727$, $p < .001$] and the within-participants variable of sentence type [$F_1(2,282) = 5.68$, $MS_e = 30,999$, $p < .01$; $F_2(2,210) = 5.95$, $MS_e = 28,620$, $p < .01$]. The effect of question type is, in fact, explained by elevated reading times in the relative clause question condition ($M = 586$, $SD = 283$). Post hoc tests of simple effects using a Bonferroni correction for multiple comparisons showed that the relative clause question condition resulted in higher reading times than both the superficial question condition ($M = 449$, $SD = 165$; $p < .01$) and the occasional superficial question condition ($M = 472$, $SD = 187$; $p < .05$).

The main effect of ambiguity is explained by elevated reading times for N1-disambiguated relative clause sentences ($M = 532$, $SD = 290$) relative to both ambiguous ($M = 490$, $SD = 232$; $p < .05$) and N2-disambiguated relative clause sentences ($M = 485$, $SD = 209$; $p < .05$).

There are two main interpretable results from analysis of this region. First, participants took longer to read the reflexive pronoun when they were asked about relative clauses on experimental trials. Since overall reading patterns for relative clause attachment types did not change along with this variable, it is difficult to argue that syntactic ambiguity resolution strategies shift as a result of this increased care. Thus, any model that could permit reading demands to influence overall reading rate could deal with this result. Second, the long reading times for N1 sentences in comparison with both N2 and ambiguous sentences could be accounted for by both the race and principled underspecification models by assuming that N2 generally wins the race or is favored when selection does take place, but this assumption could not be made for the original Traxler et al. (1998) data, nor does it adequately account for data we present below. We will pro-

pose that not all relevant interpretation of our sentences is completed during the reading of the reflexive pronoun.

Postdisambiguation Region

Figure 3 shows the data for this region, broken down by condition. These patterns are similar in some ways to those observed in the previous region: There was a difference in reading times between question type groups such that relative clause questions caused readers to slow down, and N1-disambiguated sentences were read more slowly than ambiguous sentences. However, more interestingly, in the conditions with superficial questions we replicated the main result reported by Traxler et al. (1998): Readers read both N1- and N2-disambiguated sentences more slowly than ambiguous sentences.

To confirm these observations, a 3 (sentence type) × 3 (question type) mixed ANOVA was conducted for reading times in the postdisambiguation region. The ANOVA revealed significant main effects of question type [$F_1(1,141) = 5.68$, $MS_e = 686,781$, $p < .01$; $F_2(1,105) = 4.19$, $MS_e = 740,584$, $p < .05$], sentence type [$F_1(2,282) = 9.05$, $MS_e = 117,438$, $p < .001$; $F_2(2,210) = 10.77$, $MS_e = 68,393$, $p < .001$], and most crucially, a significant interaction between question type and sentence type [$F_1(4,282) = 3.36$, $MS_e = 117,438$, $p < .05$; $F_2(4,210) = 3.93$, $MS_e = 68,393$, $p < .01$].

Simple-effects tests were performed to determine the source of the interaction. In order to determine whether question type influenced the effect of sentence type, three ANOVAs were performed by removing one level of the question type variable at a time. These analyses revealed two interactions: The first occurred when only the relative clause question condition and the superficial question condition were included [$F_1(2,188) = 5.19$, $MS_e = 47,072$, $p < .01$; $F_2(2,140) = 5.93$, $MS_e = 36,597$, $p < .005$], the second when only the relative clause condition and the occasional superficial condition were included
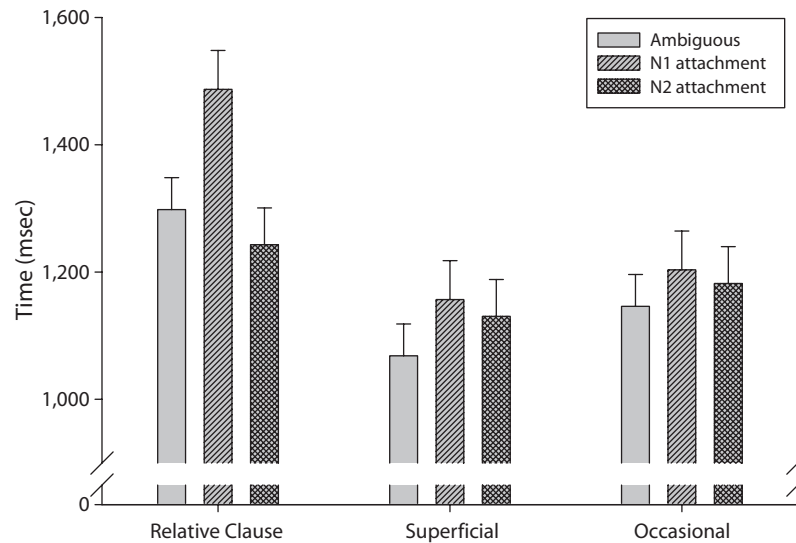
Figure 3. Reading times for the postdisambiguating region (e.g., *in public*) as a function of question type and sentence type. Error bars represent standard errors.

$[F_1(2,188) = 3.51, MS_e = 73,392, p < .05; F_2(2,140) = 5.08, MS_e = 46,475, p < .01]$. The analysis including only the two superficial question conditions revealed neither an interaction nor a main effect of question type—only a main effect of ambiguity $[F_1(2,188) = 2.51, MS_e = 51,152, p = .08; F_2(2,140) = 4.20, MS_e = 24,410, p < .05]$. The interactions can be explained as follows: When participants are regularly asked about their interpretation of the relative clause, they are more likely to spend a longer time recovering from N1 disambiguation in the previous region than from either an ambiguous or N2-resolved previous region. On the other hand, in the superficial question conditions, participants read ambiguous sentences faster than disambiguated sentences, and no reading time differences were observed for N1 versus N2 disambiguation. Table 1 presents the simple-effects analyses for each pair within each cell. Because there was no effect of question type between the two superficial question conditions, data for those conditions were pooled for the Table 1 comparisons.[4]

These results are consistent with the following conclusions. First, question type clearly influences reading times, and not merely at a gross level. Rather than showing that participants merely read more carefully when asked specific, detailed questions, this pattern of results shows something more subtle and interesting: that readers' goals may cause shifts in their syntactic ambiguity resolution strategies. When asked relative clause questions, readers spent more time in the postdisambiguation region following N1 disambiguation than they did following either ambiguous reflexive pronouns or N2 disambiguation. On the other hand, when asked superficial questions, readers were slowed down to the same extent by N1 and N2 attachments, and an ambiguity advantage was observed similar to those in studies using similar questioning methodologies (Traxler et al., 1998; van Gompel et al., 2005; van Gompel et al., 2001).

The results from the superficial question conditions are consistent with both the race model, which assumes that readers make syntactic decisions in the ambiguous condition, and the principled underspecification hypothesis, which assumes that readers tend to leave ambiguous relative clauses underspecified. However, data from the relative clause question condition fit better with the assumption of principled underspecification models that such questions are more likely to lead readers to adopt a strategy of creating attachments with increased specificity, even for ambiguous sentences, in order to be able to answer the questions. These data further suggest that when

Table 1
Planned Comparisons of the Effects of Sentence Type Within Question Type Using Bonferroni-Corrected *p* Values

Comparison	Relative Clause Questions				Superficial Questions			
	F_1	p	F_2	p	F_1	p	F_2	p
Ambiguous–N1	10.50	<.01	8.53	<.01	8.98	<.01	13.33	<.001
Ambiguous–N2	1.48	.23	2.52	.12	1.89	.14	4.13	<.05
N1–N2	14.22	<.001	15.00	<.001	<1	n.s.	<1	n.s.

Note—The degrees of freedom for the F_1 comparisons are (1,47), and the degrees of freedom for the F_2 comparisons are (1,35). *Ambiguous* refers to sentences with ambiguous relative clauses, *N1* to sentences with relative clauses disambiguated toward the first noun phrase, and *N2* to sentences with relative clauses disambiguated toward the second noun phrase.

readers must make an interpretation of ambiguous attachments, they are most likely to make an N2 interpretation. As stated in the introduction, an analysis of question-answering data will help evaluate this interpretation by revealing readers' final interpretations of sentences. We turn to these data next.

Question-Answering Data

Question-answering times were analyzed for all relative clause questions and frequent superficial questions. But first, we offer here a brief account of the response patterns for relative clause questions about ambiguous sentences, and of response accuracy for all superficial questions and for relative clause questions about the disambiguated sentence types. Consistent with a previous study that used nearly identical relative clause ambiguity materials (Traxler et al., 1998, Experiment 2), which showed a 70% N2 attachment preference for those materials during norming studies, response data reflect an overall offline preference for attaching relative clauses to N2 rather than to N1. The primary evidence for this preference is that when answering a question about their attachment of the relative clauses for ambiguous sentences, participants tended to choose to attach to N2: On average, 57.8% of their responses (SD = 22.2) favored N2. It made no difference whether the question pertained to N1 (57.6%) or N2 (58.0%) ($t <$ 1). Item-by-item preferences can be found in the Appendix.

The other evidence for an N2 preference is that participants were more likely to correctly answer questions about N2 rather than N1 sentences. In fact, accuracy patterns for relative clause questions in the disambiguated conditions revealed two main effects (with no interactions). First, participants were more accurate in answering questions about relative clauses that had been disambiguated toward N2 (M = .79, SD = .21) than about relative clauses disambiguated toward N1 (M = .73, SD =.19) [$F_1(1,47)$ = 2.44, MS_e = .08, p = .12; $F_2(1,35)$ = 4.51, MS_e = .03,

$p <$ .05]. Second, participants were more accurate when the answer to the relative clause question was "yes" (M = .79, SD = .18) rather than "no" (M = .72, SD = .19) [$F_1(1,47)$ = 3.62, MS_e = .06, p = .06; $F_2(1,35)$ = 9.44, MS_e = .02, $p <$ .01]. These accuracy data likely reveal that even though the reflexive pronoun made N2 interpretations illicit, for N1 attachment sentences some proportion of readers remembered the relative clauses as having had an N2 attachment, perhaps because their initial N2 interpretation was not fully overridden by the ultimately correct N1 interpretation.

There was no effect of sentence type on accuracy for answering superficial questions (range, 91.8%–95.1%), nor was there an effect of having to respond "yes" versus "no" (yes, 92.3%; no, 93.9%).

Figure 4 shows response time data for relative clause questions and superficial questions by sentence type. There are two obvious effects: a main effect of question type, in which relative clause questions take more time to answer, and an interaction between question and sentence type, such that relative clause ambiguity affects question response time only when the questions are about relative clause interpretation. Specifically, readers seemed to take longer to answer questions about relative clause interpretation when the attachment was left ambiguous, implying that they did not make attachments as determinately as in the unambiguous cases.

We performed a 2 × 3 × 2 mixed ANOVA in which question type (relative clause vs. superficial), sentence type, and yes-correct versus no-correct questions were used as independent variables to test effects on the time required to correctly answer questions about the experimental items. Because only one observation was collected per cell of the sentence type variable for the occasional superficial question condition, statistical analyses would be unreliable for that condition, and hence were not done. The ANOVA showed significant main effects of all three variables:
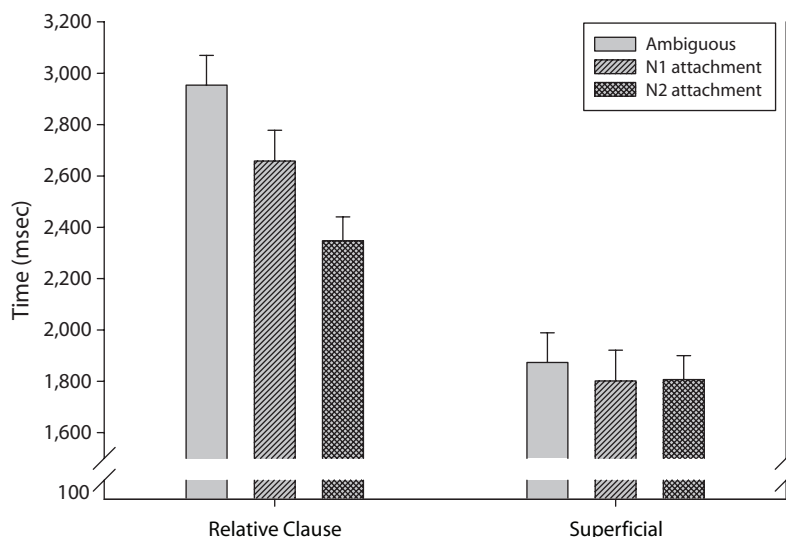

Figure 4. Question response times as a function of question type and sentence type, for correct answers only. Error bars represent standard errors.

question type [$F_1(1,94) = 30.96$, $MS_e = 3,174,336$, $p < .001$; $F_2(1,70) = 74.07$, $MS_e = 1,029,585$, $p < .001$], sentence type [$F_1(2,188) = 10.50$, $MS_e = 25,699$, $p < .001$; $F_2(2,140) = 8.04$, $MS_e = 408,071$, $p < .001$], and yes- versus no-correct questions [$F_1(1,94) = 8.68$, $MS_e = 518,435$, $p < .01$; $F_2(1,70) = 7.03$, $MS_e = 122,409$, $p = .01$]. Briefly, the main effect of yes- versus no-correct questions illustrated that participants took longer to correctly answer questions to which the answer was "no" ($M = 2,333$, $SD = 899$) rather than "yes" ($M = 2,144$, $SD = 828$).

Because the questions were so vastly different in the two question conditions discussed here, we will examine the simple effects of sentence type broken down by question type, without reference to interactions. In the relative clause question condition, there was a main effect of sentence type [$F_1(2,94) = 11.18$, $MS_e = 396,512$, $p < .001$; $F_2(2,70) = 9.04$, $MS_e = 331,569$, $p < .001$], but in the superficial question condition there was not ($Fs < 1$). Figure 4 shows that questions about the relative clauses of ambiguous sentences were answered more slowly than were questions about relative clauses in both of the disambiguated sentence types. These differences were significant in planned comparisons with Bonferroni corrections ($ps < .05$). In contrast, no differences were significant in the superficial condition ($Fs < 1$).

The implication of these results is that when a sentence was globally ambiguous, the ambiguity was not fully resolved during comprehension of the sentence. If it had been, as obligatory-attachment models predict it would be, participants would not have taken longer to answer questions about them than about disambiguated sentences. These results are consistent with principled underspecification models of attachment but are problematic for obligatory-attachment models.

GENERAL DISCUSSION

This General Discussion is organized as follows. First, we summarize the results of the experiment and briefly outline their implications for current models of sentence comprehension. Next, we discuss ways that readers' goals might be incorporated in theories of language comprehension. Before ending with a brief set of conclusions, we offer a couple of examples of how relative clauses might be parsed according to a "good-enough" approach to comprehension.

Implications of the Results

The present study yielded the following set of results: (1) an ambiguity advantage in reading times when task demands allowed for superficial interpretations of sentences, (2) a sharply different pattern of reading times when task demands required detailed interpretation of the relative clause and its attachment site, and (3) elevated response times when answering questions about sentences in which the attachment site for the relative clause was ambiguous rather than disambiguated. We draw two conclusions from these results. One is that the human sentence parsing system underspecifies attachment decisions about adjunct phrases until it is either able or pressed to make a firm decision. The other conclusion is that deciding whether to form some syntactic attachments is at least partially goal dependent. These conclusions, taken together, support models of ambiguity resolution that allow for strategic underspecification of ambiguous input, such as the good-enough (Ferreira et al., 2002; Sanford & Sturt, 2002) and construal (Frazier & Clifton, 1996) models.

These results appear to be inherently problematic for probabilistic race models. Because most of the evidence for the race model has been drawn from studies that did not report final interpretations of sentences, there is no basis for assuming that the readers who participated in those studies actually computed fully specified interpretations. In examining the present results, let us begin with the superficial question condition. In that condition, the reading time patterns were similar to those found in previous studies that seemed to provide support for race models (Traxler et al., 1998; van Gompel et al., 2005; van Gompel et al., 2001): Ambiguous sentences were read faster than disambiguated sentences. However, in the relative clause question condition, the reading time pattern was quite different, with an apparent N2 advantage appearing. Online evidence for this preference comes from the elevated reading times at the spillover region for N1 disambiguations, consistent with the 58% offline preference for N2. Race models could account for equally fast reading times in the ambiguous and N2 conditions only by assuming that N2 *always* wins the race. However, this explanation is inconsistent with the account given to explain the results in the superficial question condition.

Another piece of evidence damaging to the race model is that the pattern of question response times in the relative clause question condition suggests that participants failed to make a confident attachment of the relative clause when its attachment was ambiguous. According to the race models that have been developed to date, participants should have been equally fast to answer questions about both ambiguous sentences and sentences disambiguated by the input, for in either case, an attachment decision should have been made. The slow response times found in the ambiguous condition suggest that definitive attachments were probably not established when the sentence was read. Instead, they had to be established during the question-answering phase of the trial, resulting in long response times for the questions.

In response to these question-answering data, proponents of race models might argue that strategic postinterpretive processes explain the longer question-answering times for ambiguous sentences. The idea would be that when participants must answer a series of questions about a particular ambiguity, they may try to search for additional interpretations, even if they had initially settled on just one. We acknowledge that this is a reasonable extension of the race model, but it is not an idea that has independent motivation. In contrast, we take our inspiration from the field of text processing, which has made extensive use of question-answering paradigms to provide information about the *content* of interpretations created as text is read (see, e.g., Singer, 2006; Singer & Halldorson, 1996; Singer, Halldorson, Lear, & Andrusiak, 1992). This

work has shown that offline measures can help gauge on-line interpretation. For example, Singer (2006) reported that people who are asked to make true/false judgments to statements such as *water extinguishes fire* respond more quickly after reading texts that required that the proposi-tion be inferred than they did to texts that did not invite that same inference (Singer & Halldorson, 1996). But this "inference validation" effect was eliminated when the readers' task was merely to detect spelling errors. Thus, in an entirely independent line of work, two findings compatible with the interpretation of our data have been reported: One is that information presumably computed online leads to faster verification of that information off-line. The other is that easy offline tasks lead to shallow online processing.

We do not deny the claim made in previous studies (Traxler et al., 1998; van Gompel et al., 2005; van Gompel et al., 2001) that finding fast reading times for ambigu-ous regions poses problems for many competition-based parsing models. Neither do we claim that this finding rules out all such models as a class (cf. Green & Mitchell, 2006). Traditional constraint-based models of ambiguity resolution predict that globally ambiguous sentences with equibiased alternatives bog down processing, and there-fore slow reading times. The idea is that the two inter-pretations compete with each other, and if they are close to equally viable, the competition phase will take time to resolve. As a result, reading times for globally ambiguous sentences will be longer than for ones with a single ulti-mate interpretation. The results of the present experiment, as well as previous studies showing ambiguity advantages, are not easily explained by these models. More recently, though, Green and Mitchell argued that some competition models can accommodate the finding that ambiguous sen-tences are processed faster than unambiguous sentences—specifically, when one of the alternatives is stronger than the other. But as we stated in the introduction and as our data show, the preference for N2 attachments in the am-biguity that we studied is quite weak. We also note that some versions of constraint-based models allow for more graded resolution of syntactic ambiguity, and we suspect that competition models could be built to predict either the presence or absence of slowdown in the ambiguous region of our items. However, existing models have indeed tended to appeal to competition in cases of ambiguity.

Two issues remain to be addressed. One is how a parser might be responsive to a reader's strategies and goals, and the other is how an underspecified parse might be created. We first turn to the question of the strategy and goals in parsing.

The Importance of Goals

Theories of decision making beginning with Simon (1956) have considered the importance of goals in human reasoning. Gigerenzer, Todd, and the ABC Research Group (1999) picked up the tradition. Rational models of decision making (Chater, Oaksford, Nakisa, & Red-ington, 2003) and situation awareness (Endsley, 1995) also include an organism's goals in their models of the cognitive system. But in psycholinguistics, and especially

in work on parsing, the goals of language users have not received much attention, though one rational analysis of parsing (Chater, Crocker, & Pickering, 1998) assumed that the goal of language comprehension is to create opti-mal parses of input. However, evidence from several stud-ies (Mistler-Lachman, 1972; Townsend & Bever, 1991; Townsend, Hoover, & Bever, 2000) has shown that atten-tion to and memory for different levels of representation are susceptible to manipulations of experimental tasks. For example, in a study that partially motivated the between-participants manipulation used in this study (Mistler-Lachman, 1972), participants read sentences with four levels of ambiguity (unambiguous, lexical, structural, and logical) while performing three types of tasks: a judgment about whether the sentence made sense, a context integra-tion judgment, and a follow-up-sentence production task. Whereas the presence of ambiguities slowed responses in the latter two types of tasks, ambiguities did not slow responses to sentences when participants were asked to evaluate whether the sentences had meaning. The authors reasoned that during a simpler task, such as the mean-ingfulness judgment, one can process more shallowly by using an "anomaly search" heuristic, thereby falling short of a full interpretation.

Indeed, if the end result of comprehension is susceptible to such task manipulations, it might be that the compre-hension system does not always aim to produce an optimal parse, in the sense of creating a structure that is detailed and complete. Ferreira et al. (2002) made a more likely case for the usual goal of comprehension: to get a parse of the input that is "good enough" to generate a response given the current task. In many circumstances, the task might be to produce a response appropriate enough to sus-tain the current conversation; in other situations, the task might be to answer questions or to respond to some other type of comprehension assessment.

Sanford and colleagues (Sanford, 2002; Sturt, Sanford, Stewart, & Dawydiak, 2004) have also argued that both connectionist and symbolic models should treat depth of processing as graded. Sanford (2002) linked the notion of shallow processing to scarcity of processing resources. He argued that a system with limited processing resources should be capable of flexibly allocating them. Language comprehension is no different from other cognitive sys-tems (such as vision), and determining what information to analyze in detail and what information to treat shal-lowly in order to preserve resources must be an impor-tant aspect of understanding language. Indeed, there is evidence that comprehenders of spoken utterances devote attentional resources to the focused information in an ut-terance and process the presupposed information quite shallowly (Cutler & Fodor, 1979; Hornby, 1974). Thus, even within a single sentence, not every word is processed to the same degree: More resources are devoted to words that contribute new information to the ongoing discourse. In addition, Sturt et al. (2004) used a text change para-digm to study depth of processing during reading. Readers were instructed to detect the one word that changed across two consecutive presentations of the same short text. Sturt et al. (2004) found that the detection rate depended on the

extent to which material was linguistically focused, and they concluded that representations are only as good as the current task requires.

An Underspecified Parse

We argue that the language comprehension system can proceed in a way that not only allows underspecification, but does so strategically. When comprehenders realize (for instance) that they will not have to prove that they know whether it was the maid or the princess who scratched herself in Sentence 1, they have little reason to devote resources to this problem. This was the case in the superficial condition in the present experiment, but it was also the case in previous studies (Traxler et al., 1998; van Gompel et al., 2005; van Gompel et al., 2001) that found ambiguity advantages that were not explained with an appeal to underspecification. In this final section, we speculate about how our proposal relates to some previous language comprehension models and consider how these models could be adjusted to respond to task demands.

Construal (Frazier & Clifton, 1996) is one fairly explicit hypothesis that allows for a limited degree of underspecification of the sort we observed. Clearly, construal theory is not yet sufficiently developed to answer every question about how a limited degree of underspecification can be permitted in a parsing model, but it can help point the way toward models that may do so. According to construal, syntactic relations can be divided into primary and secondary types. Whereas primary relations (roughly, arguments) are immediately attached using modular, garden-path-like (Frazier, 1979) syntactic strategies such as late closure and minimal attachment, secondary relations (roughly, adjuncts), including relative clauses and other modifiers, initially are indeterminately "associated" with the current thematic domain, at which point other information can be called upon to resolve the association into a determinate attachment. This other information can include the plausibility of different thematic role assignments, gender appropriateness, and other types of semantic and pragmatic factors. Crucially, the extent to which the system consults information that might help resolve an ambiguity can be graded. We suggest that the goals of a reader can influence the extent to which a parser will undergo construal-like processes.

Since a relative clause is a secondary relation, in the superficial question condition of the present experiments the reader indeterminately associated the relative clause with the preceding thematic processing domain, which in our materials was the complex noun phrase. When the reflexive pronoun was then encountered, the reader attempted to find its antecedent. In the unambiguous conditions, a superficial comparison of the gender of the reflexive pronoun and its potential antecedents could inform the reader that one potential attachment was anomalous, forcing a time-consuming fully specified attachment to be made. However, in the ambiguous condition, because both candidates were possible and the questions did not require that the reader establish a single interpretation, the parser could quickly move on without fully specifying the relative clause attachment.

In contrast, in the relative clause question condition, task demands shifted the priorities of a construal parser so that the relative clause was frequently attached to one noun, even when the reflexive did not disambiguate. Reading time in the ambiguous condition was slowed when attachment had to be fully specified. In accordance with relative clause attachment preferences in English, an attachment to the most recent noun phrase was preferred (although this preference is subject to individual differences in working memory [see Swets et al., 2007, and Traxler, 2007, for more discussion] and was influenced by the materials [Gilboy, Sopena, Clifton, & Frazier, 1995; Swets et al., 2007]). When the gender of the reflexive pronoun clashed with an initial attachment based on this preference, the attachment had to be revised, further slowing down reading in the N1 attachment condition. In order to explain the slow answers to questions in the ambiguous condition, we must acknowledge that in the absence of gender disambiguation, the task demands of the relative clause question condition appear to have forced a fully specified attachment and interpretation on some trials but not on others.

Although we propose that some ambiguous relative clause attachments remain underspecified in the absence of task demands that encourage their resolution, we do not claim that all relative clause attachments remain underspecified. We note that Traxler et al. (1998) found significant effects of theta-domain on relative clause attachment (in line with the predictions of construal) and that van Gompel et al. (2001) found effects of quantifier type on attachment of prepositional phrases (resulting in satisfaction of the pragmatic demands of the quantifier). We further acknowledge that even if no decision is made about how to attach a relative clause or other adjunct, other syntactic and semantic decisions must still be made. It is interesting to speculate that the parser might underspecify an attachment decision about an adjunct in order to devote limited resources to other decisions, such as interpreting a quantifier or establishing the focus structure of the sentence (as in Sturt et al., 2004).

Parsing models that allow underspecification of ambiguous relations are certainly not new. Setting aside theories, like Marcus (1980), that posit strategic delay when ambiguity is encountered (which results in widely recognized problems—e.g., if no analysis is done, how does the parser know when to stop delaying?), several monotonic underspecification models have been developed in ways that permit less-than-complete specification of a syntactic analysis (e.g., Gorrell, 1995; Marcus, Hindle, & Fleck, 1983; Sturt & Crocker, 1996; Weinberg, 1993). These models generally assume that dominance relations, but not *immediate* dominance relations, are computed at once. Dominance relations between phrases can be open-ended and can allow for intervening material to be added to a parse. Immediate-dominance relations, on the other hand, are more fully specified. For example, a postverbal noun phrase could be integrated into a verb phrase, and thus would be parsed so that it is generally dominated by the verb phrase, but this parse would not specify whether the noun phrase is the entire verb complement or is part of a larger phrase. Subsequent work (e.g., Sturt & Crocker,

1997) elaborated on these models by positing "default" interpretations in the absence of a full commitment to a single determinate analysis.

Construal and monotonic underspecification models share something in common: flexibility in depth of processing. Of course, differences exist between the two accounts as well. Construal allows for graded levels of commitment only for nonprimary phrases, whereas Sturt and Crocker's (1996) monotonic model allows underspecification for all types of relations. It is unclear whether this model allows for task-dependent shifts in the extent to which information is processed, but construal certainly permits resource availability and task demands to impact the extent to which effort is made to progress from a mere association to a full-blown attachment.

Conclusions

The present study reports several important findings for theories of language comprehension. First, the task given to participants to assess comprehension and make sure they stayed alert during the experiment may have affected encoding processes: Easy questions may have led to underspecification, but challenging questions may have caused comprehenders to do more interpretive work. That task demands influence parsing strategies in this way implies a strategic element to parsing. Second, previously reported evidence for race-based models (e.g., Traxler et al., 1998; van Gompel et al., 2001) might have arisen in part because readers were not asked questions that demanded a specific interpretation of an attachment. Finally, the complete picture that emerges from both our reading time and question-answering performance data is that the parser's default preference is to leave adjuncts unattached if there is no basis on which to select one host over the other. That is, the comprehension system would rather create an underspecified parse than make a firm commitment that could later have to be undone, particularly when it has only a close to 50:50 chance of choosing correctly. This tendency to underspecify is consistent with the principle of "revision as last resort," which states that the parser is loath to revise a structure it has already built (for evidence, see Sturt, Pickering, Scheepers, & Crocker, 2001). By underspecifying when there is no basis on which to fully commit to an attachment decision, the parser avoids needing to revise the structure later. If it turns out that a choice must be made after the sentence has been processed, the reader may consult semantic and discourse information associated with the entire text, thus making a semantic rather than a syntactic decision.

REFERENCES

CHATER, N., CROCKER, M., & PICKERING, M. (1998). The rational analysis of inquiry: The case of parsing. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 441-468). Oxford: Oxford University Press.

CHATER, N., OAKSFORD, M., NAKISA, R., & REDINGTON, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior & Human Decision Processes*, **90**, 63-86.

CHRISTIANSON, K., HOLLINGWORTH, A., HALLIWELL, J. F., & FERREIRA, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, **42**, 368-407.

CLIFTON, C., JR., STAUB, A., & RAYNER, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain*. Amsterdam: Elsevier.

CUTLER, A., & FODOR, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, **7**, 49-59.

ENDSLEY, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, **37**, 32-64.

FERREIRA, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, **47**, 164-203.

FERREIRA, F., BAILEY, K. G. D., & FERRARO, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, **11**, 11-15.

FERREIRA, F., & HENDERSON, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 555-568.

FRAZIER, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. PhD dissertation, University of Connecticut, Storrs. Reproduced by the Indiana University Linguistics Club.

FRAZIER, L., & CLIFTON, C., JR. (1996). *Construal*. Cambridge, MA: MIT Press.

GIBSON, E., & PEARLMUTTER, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, **29**, 231-240.

GIGERENZER, G., TODD, P. M., & THE ABC RESEARCH GROUP (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

GILBOY, E., SOPENA, J.-M., CLIFTON, C., JR., & FRAZIER, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, **54**, 131-167.

GLENBERG, A. M., WILKINSON, A. C., & EPSTEIN, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, **10**, 597-602.

GORRELL, P. (1995). *Syntax and parsing*. Cambridge: Cambridge University Press.

GREEN, M. J., & MITCHELL, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory & Language*, **55**, 1-17.

HORNBY, P. A. (1974). Surface structure and presupposition. *Journal of Verbal Learning & Verbal Behavior*, **13**, 530-538.

LEWIS, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, **29**, 241-248.

MACDONALD, M. C., PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676-703.

MARCUS, M. P. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.

MARCUS, M. P., HINDLE, D., & FLECK, M. (1983). D-theory: Talking about talking about trees. In *Proceedings of the Twenty-First Annual Conference of the Association for Computational Linguistics* (pp. 129-136). Somerset, NJ: Association for Computational Linguistics.

MCRAE, K., SPIVEY-KNOWLTON, M. J., & TANENHAUS, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in online sentence comprehension. *Journal of Memory & Language*, **38**, 283-312.

MISTLER-LACHMAN, J. L. (1972). Levels of comprehension in processing of normal and ambiguous sentences. *Journal of Verbal Learning & Verbal Behavior*, **11**, 614-623.

OTERO, J., & KINTSCH, W. (1992). Failures to detect contradictions in a

text: What readers believe versus what they read. *Psychological Science*, **3**, 229-235.

REDER, L. M., & KUSBIT, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory & Language*, **30**, 385-406.

SANFORD, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind & Language*, **17**, 188-206.

SANFORD, A. J., & STURT, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, **6**, 382-386.

SIMON, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, **63**, 129-138.

SINGER, M. (2006). Verification of text ideas during reading. *Journal of Memory & Language*, **54**, 574-591.

SINGER, M., & HALLDORSON, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology*, **30**, 1-38.

SINGER, M., HALLDORSON, M., LEAR, J. C., & ANDRUSIAK, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory & Language*, **31**, 507-524.

SPIVEY, M. J., & TANENHAUS, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1521-1543.

SPIVEY-KNOWLTON, M. [J.], & SEDIVY, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, **55**, 227-267.

STURT, P., & CROCKER, M. W. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language & Cognitive Processes*, **11**, 449-494.

STURT, P., & CROCKER, M. W. (1997). Thematic monotonicity. *Journal of Psycholinguistic Research*, **26**, 297-322.

STURT, P., PICKERING, M. J., SCHEEPERS, C., & CROCKER, M. W. (2001). The preservation of structure in language comprehension: Is reanalysis the last resort? *Journal of Memory & Language*, **45**, 283-307.

STURT, P., SANFORD, A. J., STEWART, A., & DAWYDIAK, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, **11**, 882-888.

SWETS, B., DESMET, T., HAMBRICK, D. Z., & FERREIRA, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, **136**, 64-81.

TABOR, W., & TANENHAUS, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, **23**, 491-515.

TOWNSEND, D. J., & BEVER, T. G. (1991). The use of higher-level constraints in monitoring for a change in speaker demonstrates functionally distinct levels of representation in discourse comprehension. *Language & Cognitive Processes*, **6**, 49-77.

TOWNSEND, D. J., HOOVER, M., & BEVER, T. G. (2000). Word-monitoring tasks interact with levels of representation during speech comprehension. *Journal of Psycholinguistic Research*, **29**, 265-274.

TRAXLER, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, **35**, 1107-1121.

TRAXLER, M. J., PICKERING, M. J., & CLIFTON, C., JR. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory & Language*, **39**, 558-592.

VAN GOMPEL, R. P. G., PICKERING, M. J., PEARSON, J., & LIVERSEDGE, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory & Language*, **52**, 284-307.

VAN GOMPEL, R. P. G., PICKERING, M. J., & TRAXLER, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory & Language*, **45**, 225-258.

WEBER-MCROY, S., & HIRST, G. (1990). Race-based parsing and syntactic disambiguation. *Cognitive Science*, **14**, 313-353.

WEINBERG, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research*, **22**, 339-364.

NOTES

1. Green and Mitchell (2006) demonstrated that some competition models can predict fast reading times for a region that is locally fully ambiguous. However, this prediction depends on the existence of a strong bias toward one interpretation before entering this region; the ambiguous region can preserve this bias, whereas a disambiguating region sometimes forces a reversal of the bias. The materials we considered in the present research had only a mild offline bias (about 58% vs. 42%) to one of the two relevant alternatives, which may not have been enough to create the necessary strong bias prior to the ambiguous region. Furthermore, making the assumption that the processor is strongly biased toward a single interpretation in the absence of any evidence apart from (e.g.) frequency of occurrence in the language, as would have to be done for the materials used in the present experiment, seems to turn the Green and Mitchell version of a competition model into the equivalent of a single-analysis model.

2. We were made aware during the review process that a number of our items appear to make use of stereotypes in order to disambiguate attachments in the disambiguated conditions. For example, a few of these items, which were taken almost word for word from a previous study (Traxler et al., 1998), used the gender-neutral words *janitor*, *bank manager*, and *mayor* as masculine nouns to be disambiguated from more clearly feminine nouns (such as *sister*). Such stereotypes, one reviewer noted, may no longer exist among current college students. The reviewer found eight questionable items. In particular, it was noted that *mayor*, *bank manager*, *janitor*, *surgeon*, and *prostitute* all assumed certain gender stereotypes in order to block gender ambiguity. The reviewer also noted that *partner*, *assistant*, and *neighbor* do not possess clear gender roles. Although this is true, the latter words were only used in complex NPs in ambiguous conditions and were intended to be gender neutral.

In an effort to determine the validity of these items, we took a number of steps. First, we performed a Web-based norming study in which participants ($N = 30$) read simple sentences that each contained one of the nouns in the main experiment. The task was to continue a story with what happened next, beginning their follow-up sentence with a pronoun. We found that all of the items deemed questionable by the reviewer displayed gender biases to the degree appropriate for their use in the main experiment. That is, the participants had 100% agreement that the words *mayor*, *bank manager*, *janitor*, and *prostitute* were either masculine (*mayor*, *bank manager*, *janitor*) or feminine (*prostitute*). *Surgeon* had one respondent (out of 10) who used *She* in a follow-up sentence.

Next, *partner*, *assistant*, and *neighbor* were supposed to be interpretable as the same gender as the N2. The counterpart word for *partner* (continued with *He* 90% of the time) for ambiguous sentences in the main experiment was *salesman*, so the finding that participants interpreted the latter word as male (100%) posed no problem. *Neighbor* was interpreted as both male and female with nearly equal frequency, which implied that this word would maintain an ambiguity in the main experiment. However, people seemed to interpret *assistant* overwhelmingly as female (90%). This is actually a problem: Its counterpart word in ambiguous sentences was *chaplain*, which means that participants might have interpreted the purportedly ambiguous relative clause in *The assistant of the chaplain who drew attention to himself* as biased toward N2.

On the basis of these data, we reanalyzed all of the results reported in the article with the *assistant* item removed. No effects that we report herein were affected in a way that would result in a change in significance or interpretation when that one item was removed. In fact, as the Appendix shows, we found that that item (24) had an N2 preference of about 56%.

To sum up, we are confident that all items we used were appropriate for investigating the hypotheses. For one, we reused items from Traxler et al. (1998). Second, a norming study revealed that only one of the items behaved inappropriately for the original study. This item did not disambiguate the intended ambiguous sentence in the original study, as the norming data might suggest that it would. However, removing this item did not influence any of our reported results in any important manner.

3. In an attempt to rule out the potential that speed–accuracy trade-offs could have influenced reading times, we examined the reading times for each section conditional on correct answers. The data revealed that participants read more slowly when they answered questions correctly. However, with so many missing data from readers who answered all questions correctly, the statistical analyses that we performed may not have been reliable. Those tests that were computed found that the differences in reading times based on correct versus incorrect responses were not significant.

4. The reading time data from the spillover region suggest that task demands associated with different question types influenced the strategies participants used in reading the experimental items. To try to determine whether participants developed such strategies over the course of the experiment, we performed a series of analyses using several different splits of early versus late trials (first half vs. second half, first quarter vs. last three quarters, first ninth vs. last eight ninths, etc.). These analyses revealed no significant interactions between trial position and the manipulated variables of question type and sentence type. In other words, we were unable to find evidence that participants adjusted their strategies during the experimental session. One factor, however, may have obstructed our ability to detect strategy development during the experimental session: our use of a practice session in which participants were exposed to the different questioning conditions for 10 trials. The participants may have begun developing strategies for the experimental items during this session, which may have reduced the observable development that could occur once the experiment began.

APPENDIX
The 36 Experimental Items

1. The sister/brother of the actress who shot herself/himself on the balcony was under investigation. [72.2%]
   RC question: Did the sister/actress/brother get shot?
   Superficial question: Was there an investigation/award?

2. The uncle/sister of the fireman who criticized himself/herself far too often was painting the bedroom. [55.6%]
   RC question: Was the uncle/fireman/sister self-critical?
   Superficial question: Was the bedroom/kitchen being painted?

3. The partner/secretary of the salesman who amused himself/herself quite a bit was writing a letter to the editor. [44%]
   RC question: Was the partner/salesman/secretary amused?
   Superficial question: Was someone writing a letter/book?

4. The mother/father of the bride who embarrassed herself/himself at the reception was complaining to the priest. [50%]
   RC question: Was the mother/bride/father embarrassed?
   Superficial question: Did someone complain to/compliment the priest?

5. The uncle/aunt of the bishop who injured himself/herself last summer was concerned about the infection. [72.2%]
   RC question: Did the uncle/bishop/aunt get injured?
   Superficial question: Was there concern/confusion about the infection?

6. The brother/hostess of the mayor who complimented himself/herself constantly was bothered by the reporter. [55.6%]
   RC question: Did the brother/mayor/hostess like compliments?
   Superficial question: Did the reporter bother/amuse anyone?

7. The niece/nephew of the waitress who hurt herself/himself on the bicycle was angry about the incident. [56.3%]
   RC question: Did the niece/waitress/nephew get hurt?
   Superficial question: Was anyone angry/excited?

8. The sister/brother of the seamstress who entertained herself/himself most evenings was reading a book. [25.0%]
   RC question: Was the sister/seamstress/brother entertained most evenings?
   Superficial question: Did anyone have a book/video?

9. The sister/father of the schoolgirl who burned herself/himself the other day was usually very careful. [43.8%]
   RC question: Did the sister/schoolgirl/father get burned?
   Superficial question: Was someone usually careful/reckless?

10. The aunt/uncle of the nun who lost herself/himself in thought was disturbed by the noise. [56.3%]
    RC question: Was the aunt/nun/uncle lost in thought?
    Superficial question: Was there a disturbing/soothing noise?

11. The father/mother of the surgeon who made a fool of himself/herself at the party was greatly embarrassed. [50.0%]
    RC question: Did the father/surgeon/mother act like a fool?
    Superficial question: Was anyone embarrassed/proud?

APPENDIX (Continued)

12. The maid/son of the princess who scratched herself/himself in public was terribly humiliated. [68.8%]
    RC question: Did the maid/princess/son scratch in public?
    Superficial question: Was anyone humiliated/proud?

13. The great-uncle/great-aunt of the policeman who treated himself/herself after the accident was watching the news. [33.3%]
    RC question: Did the great-uncle/policeman/great-aunt need treatment?
    Superficial question: Was anyone watching television/listening to the radio?

14. The mother/father of the prostitute who killed herself/himself last summer had lived in Wales. [80.0%]
    RC question: Did the mother/prostitute/father get killed last summer?
    Superficial question: Did someone live in Wales/Spain?

15. The sister/husband of the beautician who cut herself/himself on the broken glass phoned for a doctor. [66.7%]
    RC question: Did the sister/beautician/husband get cut on broken glass?
    Superficial question: Did anyone phone for a doctor/lawyer?

16. The uncle/widow of the general who sacrificed himself/herself for the cause was the subject of the biography. [60.0%]
    RC question: Did the uncle/general/widow sacrifice himself/herself for the cause?
    Superficial question: Was a biography/poem written?

17. The grandmother/bank manager of the heiress who bankrupted herself/himself last year still made risky investments. [66.7%]
    RC question: Did the grandmother/heiress/bank manager go bankrupt?
    Superficial question: Did someone make risky/cautious investments?

18. The nephew/niece of the fisherman who drowned himself/herself in the ocean didn't know about the tricky current. [66.7%]
    RC question: Did the nephew/fisherman/niece drown?
    Superficial question: Was someone unfamiliar/very familiar with the current?

19. The neighbor/mistress of the actor who hated himself/herself for lying left town in a hurry. [68.8%]
    RC question: Was it the neighbor/actor/mistress who lied?
    Superficial question: Did someone leave/come to town in a hurry?

20. The brother/sister of the count who crippled himself/herself by falling off a horse took a long time to get over it. [62.5%]
    RC question: Was it the brother/count/sister who was crippled?
    Superficial question: Did anyone fall off a horse/donkey?

21. The son/wife of the janitor who educated himself/herself at night loved going to the theatre. [50.0%]
    RC question: Was it the son/janitor/wife who got educated at night?
    Superficial question: Did someone enjoy/hate the theatre?

22. The nephew/niece of the commander who calmed himself/herself after the tragedy was waiting for the doctor. [68.8%]
    RC question: Was it the nephew/commander/niece who calmed down?
    Superficial question: Did anyone wait for a doctor/priest?

23. The mother/father of the ballerina who found herself/himself in a lot of trouble phoned the police. [81.3%]
    RC question: Was it the mother/ballerina/father who was in a lot of trouble?
    Superficial question: Did anyone phone the police/fire department?

24. The assistant/daughter of the chaplain who drew attention to himself/herself all the time hated small children. [56.3%]
    RC question: Was it the assistant/chaplain/daughter who drew attention?
    Superficial question: Did someone hate/love children?

25. The grandmother/grandfather of the stewardess who treated herself/himself to an ice-cream cone was waiting at home. [53.3%]
    RC question: Was it the grandmother/stewardess/grandfather who had an ice-cream cone?
    Superficial question: Did anyone have ice cream/cake?

APPENDIX (Continued)

26. The grand-nephew/grand-niece of the admiral who wrote himself/herself a note admired sailors very much. [46.7%]
    RC question: Was it the grand-nephew/admiral/grand-niece who wrote the note?
    Superficial question: Did anyone admire/hate sailors?

27. The brother/sister of the butcher who invited himself/herself to the party enjoyed it quite a bit. [66.7%]
    RC question: Was it the brother/butcher/sister who invited himself/herself?
    Superficial question: Did someone have an enjoyable/horrible time?

28. The sister/nephew of the baroness who admired herself/himself an inordinate amount enjoyed all the attention. [53.3%]
    RC question: Was it the sister/baroness/nephew who admired herself/himself?
    Superficial question: Did anyone enjoy the attention they received?/ Was anyone embarrassed by the attention they received?

29. The brother/daughter of the colonel who enjoyed himself/herself a lot was out on the balcony. [66.7%]
    RC question: Was it the brother/colonel/daughter who enjoyed himself/herself?
    Superficial question: Was anyone on the lawn/balcony?

30. The maid/bodyguard of the baroness who prepared herself/himself very thoroughly came from the south. [46.7%]
    RC question: Was it the maid/baroness/bodyguard who was prepared?
    Superficial question: Was anyone from/traveling to the south?

31. The uncle/aunt of the milkman who had to support himself/herself with a low income liked to fish. [50.0%]
    RC question: Was it the uncle/milkman/aunt who was self-supportive?
    Superficial question: Did the person who liked to fish have a low/high income?

32. The niece/nephew of the duchess who looked at herself/himself in the mirror wore a blue shirt. [43.8%]
    RC question: Was it the niece/duchess/nephew who looked in the mirror?
    Superficial question: Did anyone wear a blue/green shirt?

33. The butler/maid of the actor who did not like himself/herself very much caught the armed robber. [62.5%]
    RC question: Was it the butler/actor/maid who didn't like himself/herself?
    Superficial question: Was the robber armed/unarmed?

34. The waitress/waiter of the woman who considered herself/himself a natural beauty spilled the drinks. [68.8%]
    RC question: Was it the waitress/woman/waiter who was conceited?
    Superficial question: Did anyone spill/clean anything?

35. The landlord/landlady of the businessman who had locked himself/herself in the office received a phone call. [56.3%]
    RC question: Was it the landlord/businessman/landlady who was locked up?
    Superficial question: Did anyone answer/dial the phone?

36. The daughter/son of the saleswoman who talked to herself/himself all the time walked into the room. [56.3%]
    RC question: Was it the daughter/saleswoman/son who talked to herself/himself?
    Superficial question: Did anyone enter/exit the room?

Note—Relative clause (RC) attachment preferences for the ambiguous conditions of these items are listed in brackets following the sentences. Relative clause attachment preferences are expressed by the percentage of attachments to the second noun of the complex noun phrase.