



Cognitive Science (2015) 1–33

Copyright © 2015 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12228

A Multiple-Channel Model of Task-Dependent Ambiguity Resolution in Sentence Comprehension

Pavel Logačev, Shravan Vasishth

Department of Linguistics, University of Potsdam

Received 5 April 2013; received in revised form 19 November 2014; accepted 20 November 2014

Abstract

Traxler, Pickering, and Clifton (1998) found that ambiguous sentences are read faster than their unambiguous counterparts. This so-called *ambiguity advantage* has presented a major challenge to classical theories of human sentence comprehension (parsing) because its most prominent explanation, in the form of the unrestricted race model (URM), assumes that parsing is non-deterministic. Recently, Swets, Desmet, Clifton, and Ferreira (2008) have challenged the URM. They argue that readers *strategically underspecify* the representation of ambiguous sentences to save time, unless disambiguation is required by task demands. When disambiguation is required, however, readers assign sentences full structure—and Swets et al. provide experimental evidence to this end. On the basis of their findings, they argue against the URM and in favor of a model of task-dependent sentence comprehension. We show through simulations that the Swets et al. data do not constitute evidence for task-dependent parsing because they can be explained by the URM. However, we provide decisive evidence from a German self-paced reading study consistent with Swets et al.'s general claim about task-dependent parsing. Specifically, we show that under certain conditions, ambiguous sentences can be read more slowly than their unambiguous counterparts, suggesting that the parser may create several parses, when required. Finally, we present the first quantitative model of task-driven disambiguation that subsumes the URM, and we show that it can explain both Swets et al.'s results and our findings.

Keywords: Sentence processing; Ambiguity; Parallel processing; Cognitive modeling; Unrestricted race model; URM; Underspecification; Good-enough processing

1. Introduction

How does the human sentence comprehension mechanism (hereafter, *the parser*) handle choice points when building structure incrementally? An early and influential

Correspondence should be sent to Pavel Logačev, University of Potsdam, Department of Linguistics, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany. E-mail: pavel.logacev@gmail.com

idea is the garden-path model (e.g., Frazier & Rayner, 1982), which proposes a deterministic, serial parser that operates on heuristic principles that seek to minimize structure building cost. It assumes that the parser builds a detailed syntactic representation as one reads or hears a sentence. The parser's initial decisions are based only on syntactic information, which means that it operates deterministically in the sense that, given a particular type of syntactic ambiguity, the parser always makes the same attachment decision. It is serial in the sense that only one syntactic structure is adopted if a choice exists. Over time, all these claims—seriality, determinism, complete structure building, and the priority of syntax—have come to be challenged. Constraint-based models (e.g., McRae, Spivey-Knowlton, & Tanenhaus, 1998) abandon the syntax-first assumption, allowing all kinds of information (syntax, semantics, plausibility, etc.) to be used simultaneously for making parsing decisions. The good-enough processing account (e.g., Ferreira, 2003) abandons the assumption that elaborate structural representations are built no matter what the comprehension task, but it retains all the other key features of the garden-path model. And the unrestricted race model or URM (Traxler et al., 1998; van Gompel, Pickering, & Traxler, 2000) abandons both the syntax-first assumption and determinism, allowing variability concerning which parse is built from one moment to the next.

Among all these alternatives to the classical view, the non-determinism proposal of the URM is the most provocative and interesting because it makes a very surprising prediction that is empirically attested: the ambiguity advantage. Traxler et al. (1998) found that in sentences like (1), the disambiguating word (*moustache*) was read faster in the ambiguous condition (1c) than in the unambiguous conditions (1a) and (1b).

To account for this effect, van Gompel et al. (2000) proposed a new model of ambiguity resolution, the URM. According to the URM, when the parser encounters an ambiguity, it starts building all permissible structures simultaneously, using all possible sources of information, including world knowledge. As soon as one of the structures is completed, all structure-building is terminated and the structure built first wins the race. The time taken to build a particular structure is assumed to depend on its plausibility and to vary as a function of random noise in the construction process. Since the adopted reading in each trial is the one that takes the least time to be constructed, the structure-building process is *non-deterministic* due to the influence of random noise. This means that on a given trial any one of the readings can be adopted, because any of the structure-building processes could finish faster and thus win the race.

In the sentences in (1), the parser is assumed to non-deterministically attach the relative clause *that had the moustache* either high or low as soon as it encounters the relativizer *that*. In (1a) and (1b), this attachment turns out to be wrong on some trials and the sentence has to be reanalyzed when the parser encounters *moustache*. No reanalysis is necessary in the ambiguous sentences in (1c), because it is compatible with any attachment. Here, the locus of the ambiguity advantage is at *moustache*, the point of disambiguation in (1a) and (1b).

(1)

- a. The driver of the car *that had the moustache* was pretty cool. (high attachment)
- b. The car of the driver *that had the moustache* was pretty cool. (low attachment)
- c. The son of the driver *that had the moustache* was pretty cool. (globally ambiguous)

But can this surprising prediction of the URM be explained in terms of a more classical parsing model? Swets, Desmet, Clifton, and Ferreira (2008) (SDCF henceforth) suggest that the answer is yes. In the studies concerning the ambiguity advantage (Traxler et al., 1998; van Gompel et al., 2000; van Gompel, Pickering, & Traxler, 2001; van Gompel, Pickering, Pearson, & Liversedge, 2005), reading comprehension was ensured by occasional superficial questions, which were intended to not draw attention to the attachment ambiguity. According to SDCF, the task demands did not require the parser to resolve the ambiguity in these studies, unless explicit disambiguation was provided, as in (1a) and (1b). Hence, the attachment was expected to remain underspecified in the globally ambiguous conditions. This is presumably so because not making any commitment if an ambiguity is detected is less costly (when the task does not require it) than committing to one reading and building the corresponding structure. Because participants in the above-mentioned studies were never asked questions about the relative clause attachment after reading an experimental sentence, underspecification—a version of good-enough processing—could be a feasible strategy for reducing processing effort.

To test this explanation, Swets and colleagues asked participants to read sentences like (1) and asked different kinds of questions about them. The difficulty and frequency of the questions were manipulated in a between-participants design. While 48 participants were asked questions about relative clause attachment on every experimental trial (e.g., *Did the maid/princess/son scratch in public?*), another group of 48 participants was asked superficial questions (e.g., *Was anyone humiliated/proud?*). A further group of 48 participants was asked superficial questions only occasionally (once every 12 trials).

An ambiguity advantage was found when questions were superficial. However, consistent with the underspecification hypothesis of SDCF, no such effect could be found when all questions were about the attachment of the relative clause. When such questions were asked, the globally ambiguous condition (1c) was read as fast as the low-attachment condition (1b), while the high-attachment condition (1a) was read more slowly.

(2)

- a. The son of the princess who scratched *himself* in public was terribly humiliated. (high attachment)
- b. The son of the princess who scratched *herself* in public was terribly humiliated. (low attachment)
- c. The maid of the princess who scratched *herself* in public was terribly humiliated. (globally ambiguous)

Swets and colleagues argue that the URM is unable to explain these data, because it predicts that globally ambiguous sentences should be processed faster than locally ambiguous sentences, irrespective of the kinds of questions asked. They explain the task dependence of the ambiguity advantage in terms of strategic underspecification: If questions are simple and do not require disambiguation of the sentence, the parser does not try to commit to any particular reading unless provided with a clear disambiguation cue. If the task does require disambiguation, however, the ambiguity is resolved toward the preferred reading. In this case, the parser will need the same amount of time for ambiguous sentences and those disambiguated toward the preferred reading.

In addition, Swets and colleagues found that questions concerning RC attachment were answered more slowly after ambiguous sentences than after their unambiguous counterparts. They interpret this finding as additional evidence for underspecification, because according to them, the parser sometimes (but rarely) underspecifies ambiguous sentences even when questions are about the RC. When this happens, the ambiguity has to be resolved before answering the comprehension questions, and this additional operation slows down processing. Unfortunately, Swets et al. do not discuss whether these two different behaviors (underspecification and full structure-building) are due to an inherent non-determinism of the parser, or possibly due to some participants not paying attention to the task.

However, it turns out that the SDCF data are not problematic for the URM. In order to understand why, the precise assumptions of the URM need to be examined more closely.

1.1. A re-examination of the unrestricted race model

Recall that van Gompel et al. (2000) argue that the ambiguity advantage occurs due to reanalysis cost in the disambiguated conditions. Interestingly, the ambiguity advantage effects in Traxler et al. (1998); van Gompel et al. (2001); and van Gompel, Pickering, Pearson, and Liversedge (2005), as well as Swets et al. (2008) were not found during early reading of the disambiguating word. Instead, they were found either at the post-critical region or in late reading time measures on the critical region. The URM can explain such a late effect: First, the parser non-deterministically attaches the RC as soon as possible. Later, if the disambiguating word is incompatible with the initial analysis, it may trigger reanalysis, causing a slowed reading on the disambiguating word or in the post-critical region (due to spill-over). Thus, the URM offers a strictly incremental account of the ambiguity advantage. SDCF's good-enough account, however, is not compatible with strictly incremental processing: Because disambiguation of a local ambiguity may occur on any word in the sentence, the good-enough parser must postpone the decision about whether to carry out or to underspecify RC attachment until the end of the relative clause. This is because the parser is assumed to underspecify only globally ambiguous sentences, and the end of the RC is the earliest point at which it can be sure that a sentence is globally ambiguous. Thus, Swets et al. (2008) must assume a certain amount of processing delay in order to explain that participants make use of different processing strategies in different attachment conditions.

This lack of incrementality appears to be at odds with the incrementality assumption made by most current theories of sentence processing with the notable exception of Construal (Frazier & Clifton, 1997): The parser is assumed to integrate every word into the current sentence structure as soon as it becomes available. For example, van Gompel and Pickering (2006) argue in favor of incrementality on the basis of the finding that sentences such as (3) lead to processing difficulty. This is arguably so because the string *The evidence examined* is initially interpreted as a main clause (MC), while *examined* turns out to be the verb of a reduced relative clause later. Thus, the parser is forced to revise its initial erroneous decision when it encounters the disambiguating phrase *by the lawyer*.

Because the initial decision must have been taken before the disambiguating material is processed, van Gompel and Pickering (2006) argue that this finding constitutes evidence in favor of incrementality. The same kind of argument can be made on the basis of other garden-path effects as well.

(3) The evidence examined by the lawyer turned out to be unreliable

However, the existence of garden-path effects does not provide evidence for *strict incrementality*, that is, the assumption that every word and every phrase is immediately integrated into the structure of the current sentence to the fullest extent possible. It is possible that the attachment of relative clauses (or possibly of adjuncts in general; e.g., Frazier & Clifton, 1997) can be delayed to some extent. For example, attachment of an RC may be delayed until it has been fully processed. Alternatively, RC attachment could be delayed until the head of the relative clause (the verb) has been processed. Henceforth, we will refer to the point in the sentence at which delayed RC attachment is carried out as the *attachment point*.

Such late attachment is compatible with the findings concerning the ambiguity advantage (Swets, Desmet, Clifton, & Ferreira, 2008; Traxler et al., 1998; van Gompel et al., 2001, 2005), where ambiguity advantage effects were found either on the post-critical region (which coincided with the end of the relative clause) or in late reading times measures, such as total fixation time, which are likely to have been caused by regressions originating from the post-disambiguating or later regions.

Under the assumption that RC attachment is delayed to some degree, the URM does not *need* to assume reanalysis in order to explain the ambiguity advantage. It is sufficient to assume that when the parser reaches the attachment point it tries to construct both attachments simultaneously and then terminates all structure-building as soon as the first one is constructed. Thus, the time to complete an attachment on a particular trial is equal to the completion time of the attachment process that is the fastest on this trial. Because the completion times of each process are assumed to vary from trial to trial, some attachment completion times come from one attachment process, and some from the other. It follows that the average time to complete an attachment is the mean of the shorter attachment times from all trials. By contrast, only one attachment can be made in an unambiguous sentence, so the average attachment time is the average time required to complete the relevant attachment process (high or low attachment).

Since the parser thus has a higher chance of completing the attachment relatively early in the ambiguous sentences in (1c) than in unambiguous sentences in (1a,b), the average reading time in (1c) should be shorter than in (1a,b). Hence, no reanalysis cost needs to be invoked in the unambiguous conditions. Under this version of the URM, the locus of the ambiguity advantage is also the last word of the relative clause, which is *moustache* in the case of the sentences in (1). In order to distinguish this explanation of the ambiguity advantage from the reanalysis-based explanation, we will follow the terminology of Raab (1962) (see Miller, 1982, where this work is cited) by referring this type of processing facilitation as *statistical facilitation*.

This version of the URM is compatible with Traxler et al.'s (1998) findings, as well as with the evidence for the ambiguity advantage found by van Gompel et al. (2000, 2001, 2005). Indeed, the reanalysis assumption in the original URM model is not only unnecessary but also renders the model's predictions impossible to quantify: In order to derive the predictions of a URM-with-reanalysis model, we would need to obtain estimates of the pure cost of reanalysis. Since the URM assumes non-determinism, the additional processing cost due to reanalysis in the disambiguated conditions cannot be used as an estimate of reanalysis cost. Without this information, it is impossible to derive quantitative predictions from the URM. Therefore, the statistical facilitation version of the URM is the only realization of the URM from which quantitative predictions can be derived, as we will do below.

While SDCF's account does explain the effect of question type on the occurrence of the ambiguity advantage, their claim concerning the URM's predictions does not take an important fact about the URM into consideration. Although the URM always predicts a race in case of an ambiguity, it does not necessarily predict an ambiguity advantage of the same magnitude in all possible situations. A key observation about the URM's predictions (not discussed in the URM literature as far as we are aware) is that the predicted *amount* of statistical facilitation depends on the difference between the mean reading times of the two processes involved in the race. This is so because the degree of overlap between the completion time distributions of the structure-building processes engaged in the race depends on the difference between their mean completion times (MCTs). The upper panel of Fig. 1 illustrates that a large amount of overlap between the two distributions (a small difference in means) leads to a high probability that one of the processes will finish relatively early. The lower panel of Fig. 1, on the other hand, illustrates that a small amount of overlap (a large difference in means) leads to a smaller probability of finishing early. When the overlap between the two distributions is small, the completion time distribution of the race process is largely identical to the distribution of the faster racing process.

Thus, the statistical facilitation predicted by the URM is largest when the difference between the MCTs of the racing processes is small, leading to a large overlap between the completion time distributions of the racing processes. In other words, the predicted ambiguity advantage is large when the two racing processes are equally fast or nearly equally fast and if there is a lot of variability in their completion times; when there is a

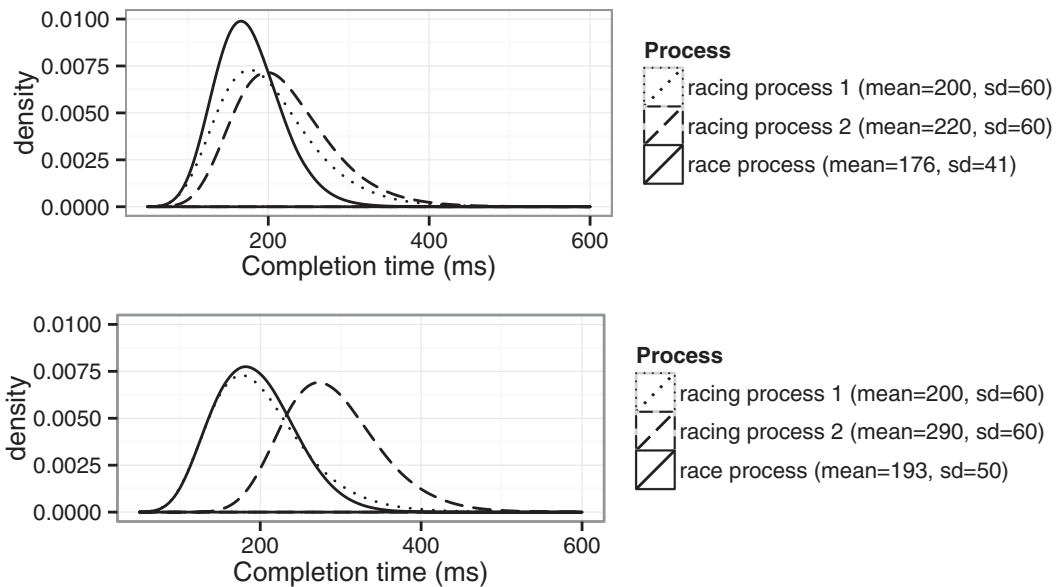


Fig. 1. Simulated completion time distributions of racing processes and the resulting race completion times: A race process has lower mean completion times if there is a large overlap between the distributions of the racing processes (upper panel); if the overlap is small (lower panel), there is little to no facilitation. The race process was simulated by repeatedly sampling one data point from each of the racing process's completion time distributions, and using the smaller of the two numbers as the completion time of the race process. Reading times of both racing processes were assumed to be log-normally distributed (e.g., Limpert et al., 2001; Ulrich & Miller, 1993), with means and SDs as provided in the legend.

large difference in between the MCTs of the processes, we may see only a very small or no ambiguity advantage at all.

Fig. 2 illustrates how the MCT of a race between two processes depends on the difference between the MCTs of the respective racing processes. To illustrate this, we simulated a race between two processes with stochastically independent completion times.

While the MCT of one process was set to 600 ms, we increased the MCT of the other racing process from 600 to 900 ms, in steps of 50 ms. Based on the finding that the standard deviation (*SD*) of a reaction time sample appears to be approximately proportional to its mean (Wagenmakers & Brown, 2007), we set the *SD* of every simulated process to 10% of its MCT. Fig. 2 demonstrates that the statistical facilitation is largest when the MCTs of the racing processes are equal, that is, when the overlap between their completion time distributions is biggest. While there is a statistical facilitation of more than 30 ms when both racing process are equally fast, its magnitude decreases as one of the racing processes becomes slower and the overlap between the completion time distributions of the racing processes decreases. For instance, when the slower of the two processes has a MCT of 750 ms, the statistical facilitation is reduced to less than 2 ms.

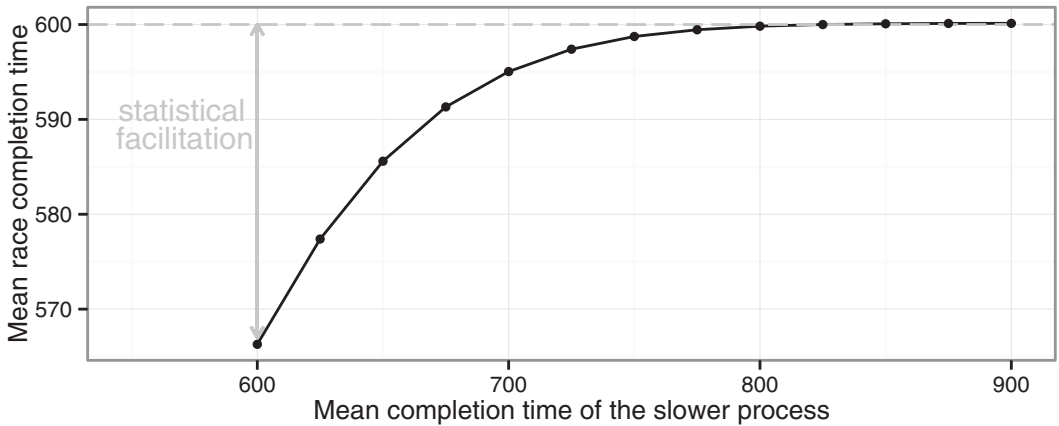


Fig. 2. Mean completion time (MCT) of a simulated race process as a function of the MCT of the slower of the two racing processes, while the MCT of the faster process remains at 600 ms. The mean race completion times are lowest when the difference between the completion times of the racing processes is small. (Simulations are based on a million samples drawn from log-normal distributions.)

1.2. An alternative explanation of the SDCF findings

This relationship between the magnitudes of the statistical facilitation and the difference between the MCTs of the racing processes has direct implications for the URM predictions concerning the magnitude of the ambiguity advantage observed by Traxler et al. (1998): The statistical facilitation account of the ambiguity advantage predicts that the ambiguity advantage should decrease with an increasing difference between the MCTs of the attachment processes.

Table 1 shows the reading times on the post-disambiguating region from SDCF. The reading times display an ambiguity advantage in the occasional and superficial questions conditions (i.e., the difference between low-attachment and ambiguous conditions reading times: 36 and 52 ms, respectively), and no ambiguity advantage in the RC questions condition (numerically, there was an ambiguity *disadvantage* of 65 ms, which was not statistically significant). Table 1 also shows that overall the post-disambiguating region was read more slowly in all attachment conditions, when questions concerned relative clause attachment. This is presumably because participants read the relative clause more carefully when they expected questions about it. Moreover, it appears that more careful reading increased the difference between reading times in high- and low-attachment sentences. High-attachment sentences were read more slowly than low-attachment sentences by 244 ms when RC questions were asked, while the reading times for such sentences differed by less than 30 ms when the questions were superficial or occasional.

Research on the effect of instructions on reading speed is consistent with the slowdown in unambiguous conditions. McConkie, Rayner, and Wilson (1973) found that participants read text passages faster if they were followed by superficial questions than if they were followed by questions requiring deeper semantic processing. In an eye-tracking

Table 1

Reading times on the post-disambiguating region, in ms (from the raw data of Swets et al., 2008). Within-subject *SEs* in parentheses (Cousineau, 2005; Morey, 2008)

	High Attachment	Low Attachment	Ambiguous
Occasional questions	1,203 (30)	1,182 (50)	1,146 (30)
Superficial questions	1,152 (27)	1,118 (24)	1,066 (23)
RC questions	1,458 (44)	1,233 (30)	1,298 (40)

experiment, Kaakinen, Hyönä, and Keenan (2002) found that readers spent more time reading a sentence when its topic was relevant to the task. In the light of these findings, slowed reading in the RC questions condition does not seem surprising. Since readers considered the relative clause more relevant when asked about its attachment after every sentence, they read the RC more carefully than when such questions were not asked. The increased difference in reading times between low- and high-attachment conditions is also not without precedent: Wotschack (2009) found an interaction between question difficulty and the effects of factors such as word frequency and word predictability in an eye-tracking experiment. For instance, the effect of word predictability was more pronounced when questions were more difficult. It appears that in Swets et al.'s experiment too, the difference between reading times in low-attachment and high-attachment conditions depended on the depth of processing.

The important point is that the size of the difference in MCTs of the two racing processes has direct implications for the magnitude of the ambiguity advantage predicted by the URM. Because the race process depends on the completion time difference between low and high attachment, the URM actually predicts a pattern very similar to SDCF's core finding: It predicts a very small ambiguity advantage in the RC questions condition. That is because the difference in reading times between high- and low-attachment conditions can be considered an estimate of the difference between the completion times of the attachment processes engaged in the race. Since this difference is larger in the RC questions condition, the URM predicts a much smaller ambiguity advantage, as illustrated in Fig. 2.

1.3. Simulation 1: URM and the effect of task-demands on reading times

1.3.1. Method

In order to establish whether SDCF's results are in principle compatible with the URM, we used the mean reading times at the post-disambiguation region in the unambiguous conditions to predict the mean reading times in ambiguous conditions. While the post-disambiguating region consisted of between one and five words (the average length was 2.8 words),¹ we assumed for our simulation that the disambiguation took place on only one of these words. Therefore, we estimated the attachment-unrelated reading time in each question condition as $1.8/2.8$ times the mean reading time in the (preferred) low-attachment sentences and assumed that the *SD* of the attachment process is 25% of the

mean reading time in that condition, based on Wagenmakers and Brown's (2007) finding of a linear relationship between mean and *SD* of reaction times. We chose to set the *SD* to an arguably plausible value of 25%, as it fit the SDCF's results well and thus allowed us to illustrate that the URM is in principle compatible with SDCF's result.²

1.3.2. Results

We simulated the predictions of the URM for ambiguous sentences in each question condition by repeatedly sampling pairs of values from two log-normal distributions (e.g., Limpert, Stahel, & Abbt, 2001; Ulrich & Miller, 1993) corresponding to the low- and high-attachment conditions and using the smaller of the two. We sampled 576 such values (corresponding to 48 subjects with 12 sentences per condition) 1 million times. The reading times for the ambiguous condition from SDCF's experiment, as well as the predictions of the fitted race model, are provided in Table 2. According to our simulation, the race model predicts an ambiguity advantage between 40 and 50 ms when questions are occasional or superficial, but a much smaller ambiguity advantage of 11 ms in the RC questions condition. An effect of this magnitude was unlikely to be detected in SDCF's experiment, and while the predicted magnitude of the ambiguity advantage differs in sign from the one obtained in the experiment, the predicted mean reading time of 1,222ms falls within the 95% confidence interval of the actual reading time in the ambiguous condition (1,220ms–1,376ms). As such, the pattern of results found by SDCF can, in principle, be explained by the race model. This makes their argument against the URM much less compelling.

1.3.3. Discussion

SDCF argue that the behavior of the parser is task-dependent. The parser underspecifies an ambiguity unless the task requires ambiguity resolution, in which case the preferred reading is chosen. Indeed, the parser's behavior *does* seem task-dependent; this is clear from the effect of question type on overall reading times in unambiguous sentences. Harder questions do indeed seem to result in "deeper" processing. However, the lack of an ambiguity advantage in the RC questions condition could simply be a consequence of deeper processing, which appears to increase the difference between the completion times of high- and low-attachment processes. Therefore, SDCF's evidence does not necessarily entail that the parser's *treatment of ambiguities* directly depends on task demands. Thus,

Table 2

Reading times for ambiguous sentences from the SDCF experiment and predictions of the URM (in ms), *SEs* in parentheses

	Ambiguous	Ambiguity Advantage	Predicted Ambiguous	Predicted Ambiguity Advantage
Occasional questions	1,146 (30)	36	1,132 (3)	50
Superficial questions	1,066 (23)	52	1,076 (3)	42
RC questions	1,298 (40)	-65	1,222 (4)	11

task demands might be modulating the MCTs in the manner discussed above, rather than parser actions per se; if so, the URM can explain the SDCF results. In other words, the SDCF findings cannot distinguish between the task-demand explanation and the URM account based on reading times alone.

However, SDCF argue that longer question-answering times in the ambiguous attachment condition provide additional evidence for underspecification. We review this claim next.

1.4. Lack of unambiguous evidence for underspecification from question-response times

Swets and colleagues found that question-answering times were significantly longer after ambiguous than after unambiguous sentences when RC questions were asked. No such differences were found when questions were superficial. SDCF interpret this finding as additional evidence for underspecification, because, according to them, participants *sometimes* (but rarely) underspecify ambiguous sentences even when questions target the RC. In such cases, RC attachment is not carried out until a question about attachment has to be answered. Their argument rests on the assumption that the content of the question cannot be compared to an underspecified representation. Therefore, the reader cannot arrive at an answer to the question without carrying out attachment first. Thus, answering questions about ambiguous attachment should require more time because carrying out attachment in the question-answering phase requires additional time, as compared to cases where a fully specified representation exists and no attachment needs to be carried out, either because sentences are unambiguous or because questions are superficial.

SDCF's account of the parser's actions predicts longer question answering times after ambiguous sentences than after unambiguous sentences, but only when questions target the RC. This is because RC attachment is carried out during reading in unambiguous sentences, but during question-answering in ambiguous sentences. Their account also predicts no such response time difference when questions are superficial. As a consequence, SDCF's explanation predicts an interaction between the factors *question type* and *attachment*. More specifically, it predicts larger differences in response times between the ambiguous and *both* unambiguous conditions when questions are about the RC as compared to superficial questions.

SDCF do not report a test for an interaction in question answering times, because "*the questions were so vastly different in the two question conditions.*" Since it is exactly that difference between questions that is assumed to drive the effect, we think it is justified to test for an interaction between the two factors. Only if question type has a larger effect on ambiguous conditions than *both* unambiguous counterparts can we consider longer question-answering times in the RC questions condition to be evidence for SDCF's account.

1.4.1. Method

We reanalyzed SDCF's question-response times with linear mixed-effects models (Baayen, 2008; Gelman & Hill, 2007; Pinheiro & Bates, 2000) using *lme4* package (Bates,

Maechler, Bolker, & Walker, 2013) in R (R Core Team, 2013). We included fixed effects of attachment and question type into the linear mixed-effects model, as well as random intercepts for participants and items. We also included random slopes for attachment (by participants and by items), question type (by items), and for the interaction between attachment and question type (by items).³

We used a treatment contrast for the factor *question type*, with superficial questions coded as 0 and RC questions as 1. For the factor *attachment*, we used two treatment contrasts with ambiguous sentences coded as 0, and the unambiguous counterparts coded as 1. Thus, we compared the question response times ambiguous to both unambiguous attachment conditions.

We conducted all analyses on log-transformed reaction times, because the Box–Cox method (Box & Cox, 1964; Venables & Ripley, 2002) suggested the logarithm as the most appropriate transformation. We excluded all response times smaller than 600 (an implausibly small value for question-response times), resulting in 14 excluded data points (i.e., 0.4% of the data). The distribution of residuals was approximately normal.

1.4.2. Results

Table 3 shows the mean question-response times in the superficial and RC questions conditions. Table 4 shows the details of our linear mixed-effect model fit: We found significantly higher answering times in RC question conditions than in superficial questions conditions ($\hat{\beta} = 0.39$, $SE = 0.06$, $t = 6.08$). There were no significant differences between the ambiguous and unambiguous attachment conditions when questions were superficial ($t = -0.91$, and $t = -0.98$).

A likelihood-ratio test comparing the model in Table 4 with a model without either interaction term revealed a significant interaction between the factors *attachment* and *question type* ($\chi^2(4) = 15.8$, $p = .003$). In the model with the main effects and both

Table 3
Question-response times from Swets et al. (2008), SEs in parentheses

	Ambiguous	High Attachment	Low Attachment
Superficial questions	1,898 (101)	1,826 (81)	1,846 (88)
RC questions	2,954 (185)	2,801 (179)	2,490 (137)

Table 4
Linear mixed-effects models coefficients, their SEs, and corresponding t -values, for the analysis of question-response times in the Swets et al. experiment

	Est. (SE)	t
RC-questions	0.39 (0.06)	6.08
High-Ambiguous	-0.02 (0.03)	-0.91
Low-Ambiguous	-0.02 (0.03)	-0.98
RC-questions \times High-Ambiguous	-0.05 (0.04)	-1.16
RC-questions \times Low-Ambiguous	-0.12 (0.04)	-3.3

interaction terms in Table 4, we found a significantly larger difference between the ambiguous and low-attachment conditions when questions were about the RC ($\hat{\beta} = -0.12$, $SE = 0.04$, $t = -3.3$), but no such effect for the difference between ambiguous and high-attachment conditions ($\hat{\beta} = -0.05$, $SE = 0.04$, $t = -1.16$).

1.4.3. Discussion

Our results show that while there is an interaction between the factors *question type* and *attachment*, its form not entirely in agreement with SDCF's predictions. That is because we found no evidence that RC questions cause a larger slowdown in question answering for ambiguous sentences as compared to their high-attachment counterparts—only the the difference between ambiguous and low-attachment conditions appears to be influenced by question type. SDCF's account, however, predicts that participants carry out RC attachment during the question answering phase in ambiguous conditions, but not in high- and low-attachment conditions. Thus, it predicts a significant effect of both interaction terms (*RC-questions* \times *High-Ambiguous* and *RC-questions* \times *Low-Ambiguous*) in Table 4. However, only one of them appears to have a significant effect.

Without any additional assumptions, the SDCF's account can only explain the absence of a significant effect of *RC-questions* \times *High-Ambiguous* as a statistical type-II error. Under the URM perspective, on the other hand, the significant effect of *RC-questions* \times *Low-Ambiguous* must be considered a statistical type-I error, because the URM does not predict any differences in question response times. Because both possibilities are compatible with the present results, Swets et al.'s question-answering data do not constitute evidence against the URM.

While we do not have enough evidence to decide between the two options, we find the second explanation preferable on the grounds of parsimony. That is because in order to explain the effect of question type on *reading times*, Swets et al. need to assume that in ambiguous sentences, readers carry out RC attachment during reading in RC questions conditions, but not in superficial conditions. However, in order to explain the predicted slowdown in question response times to RC questions about ambiguous sentences, they need to assume that readers may resort to underspecification even when questions target RC attachment. The mechanism leading to this behavior, however, adds degrees of freedom to SDCF's model which do not appear to be independently motivated.

1.5. Task-dependence of disambiguation

In sum, we have shown that both of SDCF's core findings do not provide unequivocal evidence for task-dependence of the disambiguation strategy: The apparent absence of ambiguity advantage in the RC question condition, as well as differences in question answering times are both, in principle, compatible with the URM. Thus, SDCF's data do not provide conclusive evidence for the influence of task demands on the treatment of ambiguities because, under some circumstances, the URM and the good-enough theory predict very similar patterns.

However, there is a stronger test of the influence of task demands on parsing: When the task requires *both* readings of an ambiguous sentence to be computed, the task demand account should predict an ambiguity *disadvantage*. Although this situation was not explicitly discussed by Swets et al. (2008), we believe that it is in the spirit of Swets et al.'s proposal that the parser should compute several readings of an ambiguous sentence if the task demands require it to. Thus, ambiguous sentences should be read more slowly, because computing two readings must be more costly than computing one. Importantly, an extension of SDCF's proposal must assume that parses are computed successively. In other words: The parser first attaches high, and then low, or vice versa. Since underspecification is SDCF's explanation for the ambiguity advantage, the assumption of parallel computation is incompatible with SDCF's proposal.

Importantly though, the URM can also be extended to account for the influence of task demands. To illustrate how the influence of task demands can be implemented in a parallel, obligatory attachment model, we will next present an extension of the URM the behavior of which is influenced by task demands. The critical difference between SDCF's model and ours is that the computation of several parses proceeds in parallel in our model, but serially in SDCF's model.

2. A multiple-channel model of the ambiguity resolution

Like the URM, our proposed model, the *stochastic multiple-channel model of ambiguity resolution (SMCM)* is an obligatory attachment model, which stipulates that when the processor begins processing an ambiguity, it starts building all permissible parses simultaneously. Each structure-building process can be considered to be a separate processing channel. But while the URM assumes a fixed *stopping rule* (e.g., Townsend & Colonius, 1997), according to which all structure-building terminates as soon as one permissible structure has been built, the SMCM stipulates that the stopping rule is determined by task demands: If the task does not require more than one permissible RC attachment to be constructed, the parser stops after one processing channel has terminated, that is, after one attachment has been computed. Such a system is said to be *parallel first-terminating*, following Colonius and Vorberg's (1994) terminology. If the task requires access to all available RC attachment options, the parser may choose to wait for all permissible structures to be built. Such a system is said to be *parallel exhaustive*.⁴ The SMCM stipulates that the *stopping rule* is task-dependent, and that readers, in an effort to minimize reading time, prefer a first-terminating rule, unless the task suggests that the exhaustive rule should be used. The exhaustive strategy may be preferred when the reader is aware that the sentence may have several meanings and either wants to (a) pursue both possible parses or (b) wants to select the one with the most felicitous reading. When the first-terminating stopping rule is used, the SMCM is equivalent to the URM and predicts an ambiguity advantage (cf. Fig. 3, upper panel), but when the exhaustive stopping rule is applied, the predictions reverse such that an ambiguity *disadvantage* should be observed (cf. Fig. 3, lower panel). The SMCM with an exhaustive stopping rule makes this predic-

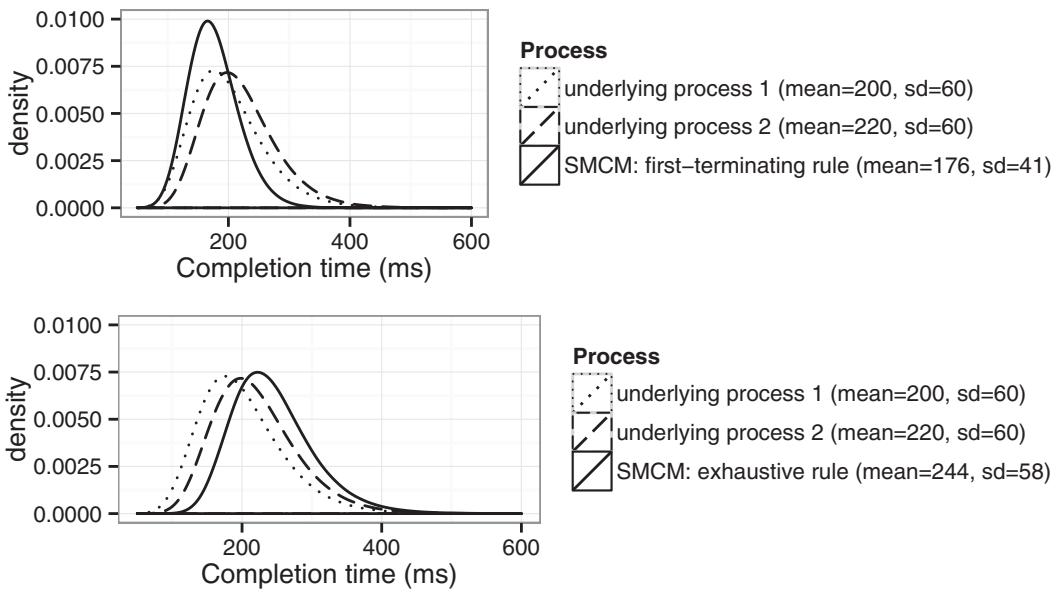


Fig. 3. Simulated completion times distributions predicted by SMCM and the underlying attachment processes. The RTs of both racing processes are assumed to be log-normally distributed, with means 200 and 180 ms and a *SD* of 60 ms. Upper panel: SMCM with a first-terminating stopping rule. The SMCM predicts shorter mean completion times (MCTs) than those for any of the underlying attachment processes. Lower panel: SMCM with an exhaustive stopping rule. The SMCM predicts longer MCTs than those for any of the underlying attachment processes.

tion because the probability that one of two attachment processes will finish relatively late is bigger than the probability that one particular process will finish late. The fact that the parser waits for both attachments to be computed in the ambiguous condition, as opposed to only one particular process in unambiguous conditions, will lead to more instances of long completion times in the former case. Thus, the mean reading time is predicted to be longer in the ambiguous condition.

Importantly, the SMCM makes the same qualitative predictions as a serial model; that is, it predicts an ambiguity disadvantage when both readings have to be computed. But while serial models are fairly unconstrained concerning the quantitative relationship between completion times in ambiguous and unambiguous conditions, the SMCM makes very precise predictions given two further assumptions:⁵

1. The completion times of the structure-building processes are *statistically independent*. This means that the speed of constructing one particular structure does not depend on the speed of any other structure-building process. In other words: The completion times of the processing channels are uncorrelated.
2. The speed of a processing channel does not depend on whether another channel is active or not. This means, for instance, that making a high attachment takes a fixed amount of time (on average), whether low attachment is permissible, or not. As Townsend and Honey (2007) point out, this assumption of *context invariance*

is the theoretical link that justifies the comparison of data in ambiguous conditions with data in unambiguous conditions.

Because Swets et al.'s findings are compatible with the URM, they are also compatible with the the SMCM employing the first-terminating stopping rule. There is reason to believe that the design of SDCF's experiment encouraged the use of the first-terminating stopping rule in all question conditions: In the superficial and the occasional questions conditions, RC attachment was not required to answer the questions correctly, which is why the parser chose the first-terminating stopping rule, in order to minimize computational effort. In the RC question conditions, on the other hand, the phrasing of the questions may have motivated participants to use the first-terminating rule: Recall that participants were asked questions such as "*Did the maid scratch in public?*" after reading sentences like (4). The readers had to respond with either "yes" or "no," while the option "*I don't know*" was not available. Asking such "yes"/"no"-questions amounts to presupposing that the reader must know the correct answer. This, in turn, is only possible if the sentence is unambiguous, because given an ambiguous sentence, a reader cannot possibly know the correct answer, since it depends on the intended RC attachment. If the RC in (4) attached high, the correct answer to the above question would be "yes," but if it attached low, the answer would be "no." Therefore, the questions may have led the participants to treat the sentences as if they were unambiguous, thus leading them to pursue a first-terminating strategy since waiting for a second parse only makes sense if sentences are ambiguous.

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

In the following, we will first test the idea of task-dependence of ambiguity resolution experimentally, and then test the qualitative predictions of SMCM.

3. Experiment

In our experiment, we asked participants to read German sentences of the form of (5), in which the RC attachment could be high, low, or ambiguous. All experimental sentences contained an NP-of-NP complex noun phrase, followed by a relative clause that either attached unambiguously to the first noun phrase, as in (5), unambiguously attached to the second noun, as in (6), or was globally ambiguous, as in (7). Disambiguating was effected by using the fact that in German, the relative pronoun indicates coindexation with a masculine or feminine noun through gender marking. A relative pronoun in a subject relative coindexed with a masculine noun is written *der*, whereas a relative pronoun in a subject relative coindexed with a feminine noun is written *die*. Thus, global ambiguity can be induced by making both nouns masculine or both nouns feminine, and unambiguous high and low attachment can be induced by making the first (respectively, second) noun match in gender with the relative pronoun. In order to counterbalance gender across conditions, there were two versions of each condition: one with a masculine, and one with a feminine relative pronoun. Recall that in English, disambiguation was

manipulated by introducing a reflexive inside the relative clause, after the relative clause verb was read; by contrast, in our experiment, attachment was disambiguated at the earliest possible point, at the relative pronoun.

- (5) *High-attachment*
- | | | | | | |
|----|----------------------------|---------------------|--------------------------|-----------------|-----|
| a. | Die Buchhalterin | des | Unternehmers, | die | ... |
| | <i>The accountant.FEM</i> | <i>the.POSS</i> | <i>entrepreneur.MASC</i> | <i>who.FEM</i> | |
| b. | Der Buchhalter | der | Unternehmerin, | der | ... |
| | <i>The accountant.MASC</i> | <i>the.POSS</i> | <i>entrepreneur.FEM</i> | <i>who.MASC</i> | |
| | ... viel goldenen | Schmuck hat, | hat momentan | Urlaub. | |
| | <i>lots golden</i> | <i>jewelry has,</i> | <i>has currently</i> | <i>holiday.</i> | |

“The accountant of the entrepreneur, who has a lot of golden jewelry, is on holiday at the moment.”

- (6) *Low-attachment*
- | | | | | | |
|----|----------------------------|---------------------|--------------------------|-----------------|-----|
| a. | Die Buchhalterin | der | Unternehmers, | die | ... |
| | <i>The accountant.MASC</i> | <i>the.POSS</i> | <i>entrepreneur.FEM</i> | <i>who.FEM</i> | |
| b. | Der Buchhalter | des | Unternehmerin, | der | ... |
| | <i>The accountant.FEM</i> | <i>the.POSS</i> | <i>entrepreneur.MASC</i> | <i>who.MASC</i> | |
| | ... viel goldenen | Schmuck hat, | hat momentan | Urlaub. | |
| | <i>lots golden</i> | <i>jewelry has,</i> | <i>has currently</i> | <i>holiday.</i> | |

“The accountant of the entrepreneur, who has a lot of golden jewelry, is on holiday at the moment.”

- (7) *Ambiguous attachment*
- | | | | | | |
|----|----------------------------|---------------------|--------------------------|-----------------|-----|
| a. | Die Buchhalterin | der | Unternehmers, | die | ... |
| | <i>The accountant.FEM</i> | <i>the.POSS</i> | <i>entrepreneur.FEM</i> | <i>who.FEM</i> | |
| b. | Der Buchhalter | des | Unternehmerin, | der | ... |
| | <i>The accountant.MASC</i> | <i>the.POSS</i> | <i>entrepreneur.MASC</i> | <i>who.MASC</i> | |
| | ... viel goldenen | Schmuck hat, | hat momentan | Urlaub. | |
| | <i>lots golden</i> | <i>jewelry has,</i> | <i>has currently</i> | <i>holiday.</i> | |

“The accountant of the entrepreneur, who has a lot of golden jewelry, is on holiday at the moment.”

After reading a sentence, participants were asked questions such as (8), in which we wanted to know if the sentence they just read “*stated or possibly implied*” a particular attachment. With questions phrased that way, the correct answer when sentences are ambiguous is always “yes,” because both attachments (high and low) are among the possible meanings of an ambiguous sentence. Thus, in order to answer the question correctly, the parser needs to construct both readings. Therefore, we would expect readers to construct both possible parses in the ambiguous condition if disambiguation is subject to task demands. It follows that SMCM as well as SDCF’s account predict an *ambiguity disadvantage* in the present experiment. Meanwhile, the URM predicts an *ambiguity advantage* or no (detectable) effect.

- (8) Wurde gerade gesagt, oder möglicherweise gemeint, dass der Buchhalter viel
 Was just now said, or possibly meant, that the accountant lots
 goldenen Schmuck hat?
 golden jewelry has.

“Did the sentence state or possibly imply that the accountant has a lot of golden jewelry?”

3.1. Method

3.1.1. Participants

Thirty-six undergraduate students from University of Potsdam, Germany, participated in exchange for course credit or 7 Euros.

3.1.2. Procedure

The task was self-paced non-cumulative word-by-word reading. Presentation and recording were done with the Linger software package, version 2.94 by Doug Rohde. At the beginning of a trial the whole sentence appeared, masked by underscores. Participants pressed the space bar to reveal the next word. As the next word appeared, the current one was masked by underscores again. The time between key-presses was recorded as the reading time for the word. Each sentence was followed by a question, which participants had to answer with “yes” or “no” by pressing the corresponding button on the keyboard.

3.1.3. Materials

Seventy fillers were mixed with thirty-six experimental items, each implementing the six sentences in (5), (6), and (7). Every experimental sentence began with a complex noun phrase involving two human nouns and was followed by a relative clause, which always consisted of five words. The relative clause mostly denoted a possessive relationship and the verb was always a form of “to have” (e.g., *had long hair, had a good sense of humor, had a cold*). The relative clause attached to either the first noun phrase (NP1), the second noun phrase (NP2), or either. Relative clause attachment was disambiguated by gender agreement between the relative pronoun and the antecedent. In order to counterbalance gender across different types of attachment (*low, high, and ambiguous*), each attachment was implemented in two sentences for every item, one with a masculine relative pronoun, and one with a feminine relative pronoun. All questions combined the proposition of either the relative clause (RC) or the MC with either NP1 or NP2. This resulted in four types of questions (RC/NP1, RC/NP2, MC/NP1, MC/NP2). Examples are provided in Table 5. The correct response was always “yes” for MC/NP1 questions, and always “no” for MC/NP2. The correct response to RC/NP1 questions was “yes” in the high-attachment condition, and “no” in the low-attachment condition. Correct responses were reversed for RC/NP2 questions. In ambiguous conditions, the correct response was always “yes” because all questions were embedded in the sentence frame “Did the sentence state, or possibly imply that ____?” (“Wurde gerade gesagt, oder möglicherweise gemeint, dass ____?”). Two-thirds of the questions were about the MC, while one-third was about the relative clause. This was done in order to avoid focusing participants’ attention on the relative clause. To make each item appear in every condition across participants, six lists were initially created. Then, in order to present every sentence with both, RC and MC questions, 12 lists were created by pairing a different subset of 12 items with questions about the relative clause.

3.2. Question norming study

We conducted a questionnaire study in order to ensure that our questions had the desired effect; that is, that they encouraged participants to construct both readings. Twenty-four undergraduate students from the University of Potsdam participated in exchange for course credit. We mixed our 36 experimental items with 36 filler sentences and asked participants to answer questions about them. All questions concerning experimental sentences in the questionnaire concerned relative clause attachment. Table 6 shows the proportions of “yes”-responses by condition. We excluded the data of three participants, because they responded incorrectly to questions about one of the unambiguous conditions in more than 80% of the cases.

If readers were to always adopt exactly one reading, the probabilities of adopting a high-attachment or a low-attachment reading, respectively, should sum to one. Thus, the probabilities of replying “yes” to either question type (RC/NP1 and RC/NP2) should sum to 1 if readers adopt exactly one reading on any given trial. Clearly, when no reading is adopted (e.g., due to attention loss), the probabilities will sum to less than one.

This appears to be so in the high- and low-attachment conditions, where the sums of the percentages of “yes”-responses are 0.95, and 0.92, respectively. This is because participants constructed the only permissible parse in these conditions. In the ambiguous condition, however, the proportion of “yes” responses to questions about NP1 is 0.80. If participants were building only one structure, we would expect the proportion of “yes” responses to questions about NP2 to be approximately 0.20. However, this proportion is 0.66, with a *SE* 0.04. Therefore, participants must have constructed both structures on at least some proportion of trials. However, the present results also suggest that our participants did not always construct both readings. If they had done so, the proportions of “yes” responses to questions about ambiguous sentences should approximately equal those of their unambiguous counterparts. That is, we would expect a proportion of

Table 5
The four different kinds of questions asked in the experiment

Did the sentence state or possibly imply that, ...	
... the accountant has a lot of golden jewelry?	(RC/NP1)
... the entrepreneur has a lot of golden jewelry?	(RC/NP2)
... the accountant is on holiday leave?	(MC/NP1)
... the entrepreneur is on holiday leave?	(MC/NP2)

Table 6
Question norming study: Proportion of “yes”-responses by attachment condition and question type. *SEs* in parentheses

	High Attachment	Low Attachment	Ambiguous
RC/NP1	0.91 (0.03)	0.10 (0.03)	0.80 (0.04)
RC/NP2	0.04 (0.02)	0.82 (0.04)	0.66 (0.04)

approximately 0.91 for *RC/NP1* questions and approximately 0.82 for *RC/NP2* questions. In spite of that, however, our data show that there is a strong tendency to compute both readings, as opposed to only one. Thus, our questions encourage participants compute both readings.

3.3. Results

3.3.1. Reading times

Table 7 provides an overview of the mean reading times for each word position in the relative clause, and Table 8 shows the reading times for the spill-over regions. We excluded the data of one participant who answered questions concerning filler sentences with a near-chance accuracy (54%), while the remaining participants performed at an accuracy above 78%. We used the *lme4* package (Bates et al., 2013) in R (R Core Team, 2013) to fit linear mixed-effects models (Baayen, 2008; Gelman & Hill, 2007; Pinheiro & Bates, 2000) to the reading time data. To determine the appropriate transformation for the dependent variable, we used the Box–Cox method (Box & Cox, 1964; Venables & Ripley, 2002). The reciprocal transformation ($1/RT$) was suggested as the most appropriate transformation for all regions we conducted analyses on. Therefore, all analyses presented here are based on reciprocally transformed reading times ($-10^5/RT$). All models included fixed effects of attachment using treatment contrasts with the ambiguous condition as a baseline. We also included random intercepts for participants and items, as well as by-participant and by-item random slopes.⁶ Outlier removal was performed using a variant of the technique recommended by Baayen and Milin (2010): We iteratively removed the data point corresponding to the largest outlying residual until the model's residuals appeared approximately normal. The rationale behind our exclusion criterion

Table 7

Mean reading times in ms for all word positions in the relative clause. Within-subject *SEs* in parentheses (Cousineau, 2005; Morey, 2008)

	Relative Pronoun	Adverb	Adjective	Noun	RC Verb
High attachment	543 (17)	481 (12)	529 (15)	528 (13)	622 (24)
Low attachment	541 (15)	497 (14)	516 (23)	548 (17)	637 (30)
Ambiguous	562 (18)	486 (14)	512 (15)	544 (15)	765 (41)

Table 8

Mean reading times in ms for the spill-over regions. Within-subject *SEs* in parentheses (Cousineau, 2005; Morey, 2008)

	RC Verb +1	RC Verb +2	RC Verb +3
High attachment	492 (14)	459 (11)	693 (31)
Low attachment	513 (17)	516 (25)	740 (36)
Ambiguous	524 (17)	503 (17)	773 (44)

was to fulfill the assumption of normality of residuals while making maximum use of the information in the data by excluding as few data points as possible and while avoiding a *one-size-fits-all* exclusion criterion, which will lead to more missing data for slower participants. For each of the reported models we had to exclude ten data points or less. For the analysis at the RC verb (the critical region), we excluded two values, both smaller than 182 ms. For all six models taken together, we excluded 43 values ranging from 127 to 2179 ms, with a median of 188 ms. In all models presented, $|t| > 2$ and $|z| > 2$ correspond to a significant effect at a significance level of .05.

Tables 9 and 10 provide the details of the analysis. We found no significant effects of attachment in reading time at the relative pronoun ($|ts| \leq 1.01$), or the noun phrase consisting of the three following words ($|ts| \leq 0.06$). There were no significant differences in reading times at any of the three words making up the noun phrase either (all $|ts| < 1.5$). However, we found a significant difference at the verb: It was read more slowly in ambiguous sentences compared to low-attachment sentences ($\hat{\beta} = -12.44$, $SE = 5.53$, $t = -2.25$) and high-attachment sentences ($\hat{\beta} = -13.96$, $SE = 4.47$, $t = -3.12$). Furthermore, the ambiguous condition was read more slowly than the high-attachment condition at the second word after the verb ($\hat{\beta} = -12.37$, $SE = 3.88$, $t = -3.19$) and the word after that ($\hat{\beta} = -9.74$, $SE = 4.63$, $t = -2.1$).⁷

3.3.2. Accuracy

Table 11 provides an overview of the mean percentages of correct question responses. Here, too, we used the *lme4* package (Bates et al., 2013) in R (R Core Team, 2013) to fit generalized linear mixed-effects models assuming a logit link function to the response data. Correct responses were coded as 1, incorrect responses as 0. All models included random intercepts for participants and items, as well as by-item and by-participant random slopes for all fixed effects, but no correlations between random intercepts and slopes. We analyzed the response accuracy for MC questions and for RC questions separately.

A multilevel model fit to MC questions accuracy data (cf. Table 12) revealed a significant effect of attachment type indicating higher accuracy in high-attachment conditions than in ambiguous conditions ($z = 2.67$). We found no significant difference between low-attachment and ambiguous conditions. Another model, which was fit to RC questions accuracy data (cf. Table 13) revealed a significantly higher accuracy in both unambiguous

Table 9

Linear mixed-effects models coefficients, their *SEs*, and corresponding *t*-values, for the analyses of reading times at the regions relative pronoun, noun phrase (adverb + adjective + noun), and RC verb

	Relative Pronoun		Noun Phrase		Verb	
	Est. (<i>SE</i>)	<i>t</i>	Est. (<i>SE</i>)	<i>t</i>	Est. (<i>SE</i>)	<i>t</i>
High-ambiguous	-4.25 (4.21)	-1.01	0 (1.06)	0	-13.96 (4.47)	-3.12
Low-ambiguous	-0.91 (4.32)	-0.21	0.07 (1.21)	0.06	-12.44 (5.53)	-2.25

Table 10

Linear mixed-effects models coefficients, their SEs, and corresponding t-values, for the analyses of reading times at the three spill-over regions after the RC verb

	Verb + 1		Verb + 2		Verb + 3	
	Est. (SE)	t	Est. (SE)	t	Est. (SE)	t
High-ambiguous	-5.62 (3.9)	-1.44	-12.37 (3.88)	-3.19	-9.74 (4.63)	-2.1
Low-ambiguous	0.03 (4.01)	0.01	-3.48 (4.83)	-0.72	-6.6 (5.48)	-1.2

Table 11

Accuracy by attachment and type of question. Within-subject SDs for proportions in parentheses (Cousineau, 2005; Morey, 2008). For RC questions in the ambiguous conditions, only “yes”-responses were considered correct

	MC Questions Accuracy	RC Questions Accuracy
High attachment	0.96 (0.01, $n = 280$)	0.87 (0.03, $n = 140$)
Low attachment	0.85 (0.03, $n = 280$)	0.79 (0.04, $n = 140$)
Ambiguous	0.86 (0.03, $n = 280$)	0.42 (0.05, $n = 140$)

than in both ambiguous conditions ($z = 4.97$, and $z = 6.05$) indicating higher accuracy for NP1 attachment conditions. Importantly, the low proportion of “yes” responses to RC questions in the ambiguous conditions was due to a low proportion of “yes”-responses to *both* types of questions: For questions about attachment to NP1, the proportion was 0.49 ($SE = 0.07$, $n = 70$), and for questions about attachment to NP2 it was 0.36 ($SE = 0.07$, $n = 70$).

3.4. Discussion

To summarize our main findings, relative clauses with ambiguous attachment, such as (7), were read more slowly than relative clauses that unambiguously attached either high or low, such as (5) or (6). This effect occurred at the verb, which was also the last word of the relative clause. This slowdown suggests that both structures, high and low attachment, were computed in ambiguous conditions. Most important, the above-mentioned slowdown at the RC verb in the ambiguous condition is unexpected under the URM account of ambiguity resolution. Its presence suggests that, at least under some task demands, there can be an ambiguity *disadvantage*. Interestingly, this effect did not occur until the last word of the relative clause, although disambiguation happened at the first word. The good-enough parsing account proposed by Swets et al. (2008) as well as the SMCM agree very well with our results, because they assume that the parsing strategy is subject to task demands. Under these accounts, the parser computes both meanings if the questions suggest that both meanings of an ambiguity should be computed. Because computing both meanings requires more time than computing just one, reading slows down at the verb. Furthermore, the location of the slowdown suggests that the relative clause attachment operation was delayed and took place at end of the relative clause (coincided

Table 12

Generalized linear mixed-effects models coefficients, their SEs, and corresponding *z*-values for the analysis of the percentage of correct answers to MC questions

	Est. (<i>SE</i>)	<i>z</i> value
High-ambiguous	2.18 (0.82)	2.67
Low-ambiguous	−0.05 (0.3)	−0.16

Table 13

Generalized linear mixed-effects models coefficients, their SEs, and corresponding *z*-values for the analysis of the percentage of correct answers to RC questions

	Est. (<i>SE</i>)	<i>z</i> value
High-ambiguous	2.99 (0.6)	4.97
Low-ambiguous	1.72 (0.28)	6.05

with the position of the relative clause verb), which is compatible with the SMCM as well as with the good-enough account. The original formulation of the URM, however, cannot explain this delay without additional assumptions.

In addition to the slowdown at the verb, we found a speedup in high attachment (relative to ambiguous sentences) one word after the verb (on positions verb+2 and verb+3), but no significant difference between high-attachment and ambiguous sentences. There are two possible explanations for this speedup. It is possible that faster reading in the high-attachment condition reflects a delayed effect of a high-attachment preference in German (Hemforth, Konieczny, & Scheepers, 2000). However, the fact that there is no such difference at the word following the relative clause renders this explanation unlikely. Thus, we believe that this effect is more likely to reflect the retrieval of the sentence subject (Lewis & Vasishth, 2005). We will assume that attaching a relative clause to a noun phrase requires its retrieval from memory, that is, reactivation of its memory trace. According to Vasishth and Lewis (2006), retrieving a constituent increases the strength of a memory trace and thus facilitates later retrievals. Thus, reading was facilitated in the high-attachment condition because only the subject noun phrase had been retrieved earlier, during the process of RC attachment, rendering it the most active noun phrase in memory, and thus facilitating its later retrieval. In the ambiguous condition, however, both noun phrases are retrieved for the purposes of RC attachment, leading to similar activation levels for both. In the low-attachment condition, only the embedded noun phrase is activated, and retrieval of the subject noun phrase is not facilitated.

In addition to the effects, we found in reading times that participants had significantly higher question-response accuracy in the MC in the high-attachment condition than in the two other conditions. A similar pattern was found for relative clause questions. These findings suggest that it may be harder to maintain a memory representation of two clauses with different subjects than to represent two clauses sharing one subject. We also found that the proportion of “yes” responses to RC attachment questions in ambiguous

sentences (i.e., 42%) was unexpectedly low, given that the slowed reading of the ambiguous relative clause suggests that both RC attachments are computed. If both parses (high and low attachment) had been maintained until the question-answering phase of the trial, the correct answer should have been “yes” in all attachment conditions. One possible explanation is that the two interpretations of the sentence somehow interfere with each other during question-answering. That is, the presence of two structures may preclude either from being retrieved during the question-answering, resulting in an incorrect “no”-answer. An alternative explanation for this surprising finding is that participants do compute both attachments of the relative clause in accordance with the hypothesized task requirements, but retain only one of the structures. A possible reason to do so is that the parser may not be able to maintain more than one reading simultaneously, or in other words that parsing is serial (e.g., Frazier, 1987; Lewis, 2000).

So why is the proportion of “yes” responses substantially higher in the question norming study than in the self-paced reading experiment? One possibility is that participants did not maintain more than one structure in our on-line experiment because the self-paced reading task taxes memory resources much more than natural reading. In the questionnaire study, however, they were able to maintain more than one structure, because they read in a more natural manner.

An alternative explanation is that while the parser never maintains more than one interpretation of a sentence, participants may have sometimes re-read sentences to check whether they had other possible meanings. This is a plausible explanation because re-reading of the target sentence was possible in our question norming study, but not in the self-paced reading experiment. Thus, when the initial disambiguation of the target sentence did not match the question, participants may have re-read the sentence to check for an alternative meaning.

In sum, our results suggest that readers can build two attachments (hence the slowdown in the ambiguous conditions), but that they possibly discard one of them. When taken together with previous findings, they show that parsing is susceptible to task demands. According to the SMCM, when question difficulty is low, processing is relatively shallow, and so structure-building terminates as soon as one parse is built—for shallow processing any one attachment will suffice. When question difficulty is high, the parser waits for both structures to be built, and then possibly selects the more plausible one. Such a mechanism predicts an ambiguity advantage in the first case, and an ambiguity disadvantage in the second.

SDCF’s theory can predict the same pattern, but for somewhat different reasons. While the ambiguity advantage is explained by strategic underspecification, an ambiguity disadvantage would need to be explained by the assumption that the parser builds two structures sequentially, when questions are difficult. SDCF would need to assume sequential structure-building because by rejecting the race account of the ambiguity advantage, they also reject the mechanism of simultaneous structure-building.

Hence, the crucial differences between the SMCM and SDCF’s theory lie in (i) the explanation for the ambiguity advantage (underspecification vs. race) and (ii) the explanation for the ambiguity disadvantage (two successive attachment operations vs. waiting for

the second of two concurrent attachment operations to finish). In spite of different mechanisms, both theories make the same qualitative predictions. However, the SMCM also makes quantitative predictions concerning the reading time in the ambiguous conditions on the basis of the reading times in the unambiguous conditions. Unfortunately, the quantitative predictions of the good-enough account are much less clear. This is so because they depend on the duration of an attachment operation, which is unknown.⁸ In the following section, we will test the quantitative predictions of the SMCM.

3.5. *Simulation 2: Testing the predictions of SMCM*

The predictions of the SMCM for our experiment are more straightforward to derive than the predictions of a serial model because we do not need an estimate of the time required by an additional attachment operation. It follows from SMCM's context invariance assumption that the variability of attachment time for high- and low-attachment processes is equal in ambiguous and unambiguous attachment conditions. Although factors such as lexical processing must undoubtedly contribute some additional variability in reading times (RT henceforth) across conditions, we expect such influences to be minor, because the verb in all items was short high-frequency word ("hat" or "hatte"). Thus, we assume that the amount of attachment-unrelated variability in RT is negligible, and therefore make the simplifying assumption that for any given participant the amount of processing time contributed by attachment-unrelated factors is constant. We therefore used the RT variability in the low- and high-attachment conditions as estimates of the attachment completion time variability of the low- and high-attachment processes, respectively. Under this simplifying assumption, we can predict the reading time in the ambiguous conditions from the reading times in the unambiguous conditions in order to examine how well the SMCM performs in the light of the evidence.

3.5.1. *Method*

In order to simulate the predictions of the SMCM with an exhaustive stopping rule, we used the data in the unambiguous conditions to estimate the completion time distributions for the high- and low-attachment processes for each participant. Based on these estimates, we repeatedly generated samples of RTs predicted by SMCM for each participant. This procedure allowed us to generate a prediction for the mean RT in the ambiguous condition, as well as to quantify our uncertainty about the prediction. Quantifying this uncertainty is important because our predictions were generated on the basis of the reading times in unambiguous conditions, which are subject to the sampling error. On the basis of the obtained variability of our predicted mean RT, we were able to determine a 95% confidence interval for our prediction.

In order to obtain a prediction, we proceeded in the following steps: For each participant and every unambiguous condition, we used the reading times at verb to estimate the parameters of a log-normal distribution representing that participant's attachment completion times in that condition. We then repeated the following steps 100,000 times:

1. For each participant, we randomly drew 12 pairs of reading times from the above-mentioned log-normal distributions. In each pair, one value was from the high-attachment distribution, and the other from the low-attachment distribution.
2. We simulated the predictions of the SMCM with an *exhaustive* stopping rule by averaging the *maxima* of the 12 pairs to obtain a bootstrap sample for one participant.
3. We averaged the bootstrap samples for all participants to obtain one bootstrap sample for the entire group.

We repeated the above steps in order to estimate the mean reading time predicted for the ambiguous condition and a 95% confidence interval based on variability in the predictions (due to the variability in the data in the unambiguous conditions). This technique (repeatedly sampling from a distribution in order to estimate the variability of a statistic of interest) is known as *stratified* or *blocked bootstrap resampling* (Hesterberg, Moore, Clipson, & Epstein, 2005; Wehrens, Putter, & Buydens, 2000). We also simulated the predictions of the SMCM with a first-terminating stopping rule by averaging over the minima instead of the maxima of the 12 pairs of reading times. The results of this simulation are shown in Fig. 4.

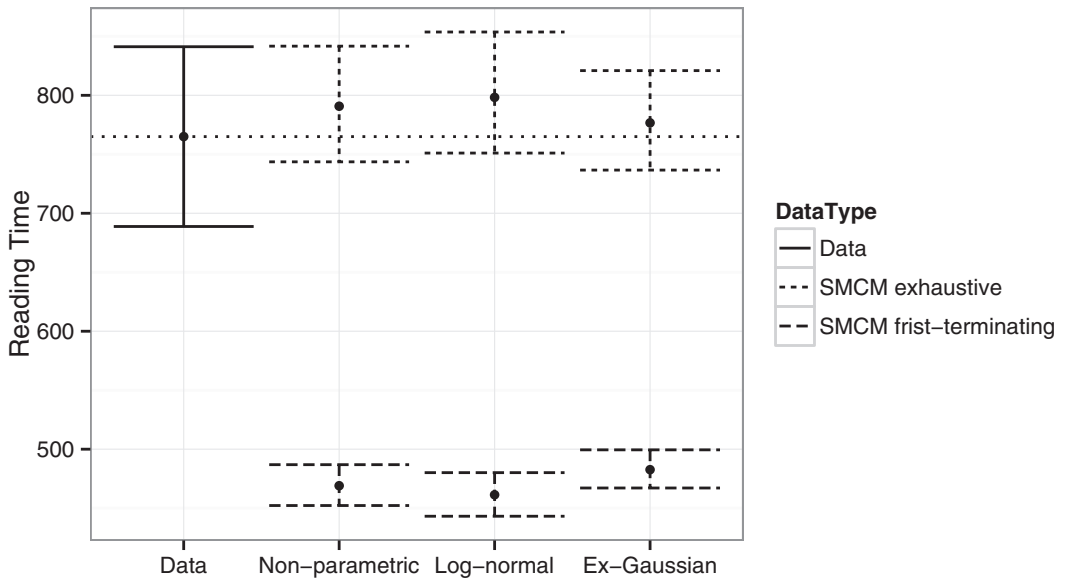


Fig. 4. Mean reading times in the ambiguous condition and mean reading times predicted by SMCM and the associated 95% confidence intervals. On the left, the empirical mean reading time at the verb in the ambiguous condition and its associated 95% confidence interval. On the right, “Non-parametric,” “Log-normal,” and “Ex-Gaussian” show mean reading times predicted by the SMCM (bold points) and the associated 95% confidence intervals obtained by bootstrap resampling based on the respective distributional assumptions for reading times in the unambiguous conditions.

Moreover, we carried out the procedure under two other sets of assumptions, to verify that our results do not depend on the assumption that RTs are log-normally distributed: (i) non-parametric bootstrapping, and (ii) parametric bootstrapping an ex-Gaussian distribution.⁹ For non-parametric bootstrapping, we sampled (with replacement) from the reading times in the unambiguous conditions and for parametric bootstrapping under the assumption that RTs in the unambiguous conditions are distributed according to an ex-Gaussian distribution. We used an iterative algorithm (Nelder & Mead, 1965) in R (R Core Team, 2013) in order to find distribution parameters maximizing the likelihood of the data in each condition (Lacouture & Cousineau, 2008; Myung, 2003).

3.5.2. Results

Fig. 4 shows the mean predicted reading times obtained via bootstrap resampling, along with 95% confidence intervals and the mean reading time for ambiguous sentences obtained in the experiment. According to our simulations, the SMCM with an exhaustive stopping rule predicted the reading time for ambiguous conditions to lie between 737 and 854 ms, corresponding to a predicted ambiguity disadvantage between 100 and 217 ms. The observed slowdown in the data was 128 ms. Thus, the empirical mean of the ambiguous condition reading times was within the 95% confidence intervals for predicted reading times in all simulations. Consequently, the empirical mean did not significantly deviate from the predicted mean reading time (all $ps > .1$, two-tailed). The SMCM with a first-terminating stopping rule (which we have included only for illustration) predicted the reading time for ambiguous conditions to lie between 443 and 499 ms, corresponding to an ambiguity *advantage* 194 and 138 ms.

3.5.3. Discussion

Our simulation results demonstrate that the empirical mean reading time in the ambiguous condition does not significantly differ from SMCM's predictions. This result held true under different distributional assumptions, which produced almost identical predictions and importantly, the reading time data from the ambiguous condition was not used in parameter estimation. Our result does not constitute evidence against good-enough theory, because successive computation of two attachments is compatible with slowdowns of any magnitude in the ambiguous condition. The results of the simulation demonstrate however, that the SMCM makes more constrained predictions than a serial model, without additional assumptions and without free parameters, and can thus be more easily tested.

3.6. General discussion

We have provided empirical evidence for the task-dependence of ambiguity resolution. In addition, we presented the SMCM, a new quantitative model that can explain the effect in terms of a non-deterministic model of parsing which assumes multiple channels and a stopping rule. This model can be seen as a refinement of the SDCF idea and is an improvement over the underspecification proposal because it makes precise assumptions

about the timing of the attachment process, and because it makes precise predictions about the relationship between the attachment times in ambiguous and unambiguous conditions.

The SMCM has one free parameter, namely the kind of stopping rule used by the parser. We assume that the stopping rule selection depends on the type of questions participants are asked, with questions in the present experiment causing the parser to use an exhaustive stopping rule. We furthermore assume that all participants use the same stopping rule on all trials. Clearly, it is possible to imagine that the same task may prompt different people to make use of different stopping rules at different times, as they may differ in their perception of task demands. Furthermore, different participants may decide to start using an exhaustive stopping rule at different points in the experiment, while others may use a first-terminating stopping rule throughout. As a result, some tasks may result in the usage of an exhaustive stopping rule on only a certain proportion of trials. Under such circumstances, the exact predictions of the SMCM will depend on this proportion, which would constitute an additional free parameter. Therefore, such a situation would not constitute a strong test of the SMCM unless we have an independent way of estimating this proportion, because this free parameter would enable the SMCM to explain a whole (restricted) range of reading times for ambiguous sentences with the right proportion parameter. However, the present data do not provide any indication that such a parameter may have to be used.

Importantly, although the SMCM is a parallel processing model with independent channels, it does not necessarily keep all the parses that have been computed. This means that it is not parallel in the sense that several parses are maintained (e.g., Lewis, 2000). We assume that the SMCM parser discards all structures except one, because of the low proportion of “yes” responses to RC questions in our experiment. This is because a “no” response to a question about a particular attachment indicates that an interpretation with this attachment was not in storage at the time the question was answered. Thus, a low proportion of “yes” responses suggests a high number of trials on which the reading asked about was not in storage. We assume that this is because the other reading was selected on that trial. There are several possible reasons for the parser to retain only one parse. One possibility is that the number of stored structures is subject to task-demands as well. While our task did modify the parser’s stopping behavior, a somewhat different task may be required to modify the parser’s storage strategy. For instance, participants may have been more inclined to store multiple interpretations of a sentence if they had been provided with feedback as to the correctness of their answer. Another possibility is that the strategy of selecting only one parse when several are available is intrinsic to the parsing algorithm and non-modifiable by task demands. Structures may be discarded due to memory limitations, or because the parser’s architecture forces it to take a decision whenever several options are simultaneously available (e.g., Frazier, 1987).

Theories of sentence processing diverge on two main issues: (i) the number of analyses maintained by the parser at any given moment, and (ii) the algorithm underlying potential disambiguation. The timing of the processes which generate different parses is rarely addressed. Only two theories make explicit claims about the timing of these processes:

the garden-path theory (Frazier, 1979 1987) and the URM (van Gompel et al., 2000). Both theories assume that analyses are created simultaneously. For instance, according to the garden-path theory, a deterministic race is the mechanism underlying the parsing principles minimal attachment and late closure. When facing a choice between two structures, the parser always adopts minimal structure because less time is required to build it. However, unlike the URM, the garden-path theory assumes no stochasticity (non-determinism) in the structure-building process, which is why the minimal attachment structure *always* takes less time.

Interestingly, *cue-based parsing* may make similar predictions when faced with ambiguity. Lewis, Vasishth, and Van Dyke (2006) assume that in order to establish a dependency between two co-dependent elements like subjects and verbs, the first-occurring co-dependent needs to be retrieved from memory. In post-nominal relative clauses, for example, RC attachment requires the parser to retrieve the noun phrase to which the RC attaches when it encounters the verb. This is because that noun phrase is an argument of the RC verb, and retrieval is assumed to precede dependency resolution. If only one reading needs to be computed, only one noun phrase needs to be retrieved. This operation may be carried out more quickly in the ambiguous condition than in an unambiguous condition because two candidate noun phrases are available in memory in the former case, as opposed to just one in the latter—and retrieval of any one of two noun phrases must finish faster on average due to a higher probability of the search finishing early.

Thus, while the exact nature of the operations engaged in a race is left open in the URM as well as the SMCM, an integration of cue-based parsing and SMCM could fill in this gap by assuming that at least a part of the computations performed simultaneously is due to search in content-addressable memory (Martin & McElree, 2009; McElree, 2000). We leave this question for future research, but fortunately, we are able to derive quantitative predictions even without the knowledge of the exact mechanism of these processes, because the reading times in the unambiguous conditions provide us with estimates of their durations.

3.7. Conclusion

We demonstrated through simulation that Swets et al. (2008) do not provide conclusive evidence for the assumption that the parser's handling of ambiguities depends on task demands. We then provided more compelling experimental evidence for the Swets et al. claim by demonstrating that given a sufficient amount of task difficulty, sentences with attachment ambiguities can be read more slowly than unambiguous sentences.

We then presented a new model, the SMCM of ambiguity resolution as an extension of van Gompel et al.'s (2000) URM. Finally, we demonstrated that in addition to being able to account for previous findings (Traxler et al., 1998; van Gompel et al., 2000, 2001, 2005) the SMCM makes more constrained predictions than a model which assumes sequential attachment. The SMCM is, to the best of our knowledge, the first quantitative model of task-dependent parsing.

Acknowledgments

We would like to thank Roger Levy, Bruno Nicenboim, Benjamin Swets, Klinton Bicknell, and one anonymous reviewer for insightful comments on earlier drafts of the manuscript.

Notes

1. Many thanks to Benjamin Swets for kindly providing us with the stimuli and data of the original study.
2. We could not use SDCF's raw data to estimate the within-participant variability in attachment completion times due to the fact that their items were not matched for length of the post-disambiguating region as well as the length and frequency of the words therein. The resulting between-items variability would have led to overestimates of the variability in the completion times the attachment process. Furthermore, this problem is exacerbated by Wotschack's (2009) finding of an interaction between lexical variables and task difficulty. This effect would selectively increase estimates of variability in the RC question conditions and thus lead to underestimates of the predicted reading time in the ambiguous condition.
3. We did not include correlations between random intercepts and random slopes, because some models produced pathological estimates of such correlations (i.e., 1 or -1) and because according to Barr, Levy, Scheepers, and Tily (2013), models without such correlation parameters do not significantly differ from maximal models in terms of controlling the Type-I and Type-II error rates.
4. Clearly, other stopping rules are theoretically possible. For example, the parser may wait for a fixed number of processes to finish. However, for our present purposes we distinguish only between the first terminating and the exhaustive stopping rules.
5. These assumptions also render it a parallel model with independent processing channels (Townsend & Ashby, 1983).
6. We did not include correlations between random intercepts and random slopes, because some models produced pathological estimates of such correlations (i.e., 1 or -1) and because according to Barr et al. (2013) models without random correlations do not significantly differ from maximal models in terms of controlling the Type-I and Type-II error rates.
7. The pattern of results remained when all RTs above 3000ms were excluded: Ambiguous sentences were read more slowly than high-attachment sentences ($t = -2.7$) at the verb, as well on the second and third word following it ($t = -3.19$, and $t = -2.1$). The slowdown for ambiguous sentences compared to low-attachment sentences at the verb was marginally significant ($t = -1.92$).
8. All we can assert about an attachment operation in our experiment is that its duration is estimated to be between 0 and 622 ms (the reading time in the faster one of

the unambiguous conditions). Although a part of these 622 ms must be due to processes such as word recognition, and pressing a button, and RC attachment is therefore likely to require significantly less time than 622 ms, we have no way of estimating its duration. Therefore, the good-enough model is compatible with any ambiguity disadvantage of less than 622 ms. The model's quantitative predictions are therefore fairly unconstrained and no meaningful quantitative predictions can be derived.

9. Both distributions, the log-normal and the ex-Gaussian, have been used as descriptive models of reaction time (e.g., Ulrich & Miller, 1993; Van Zandt, 2002).

References

- Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge, England: Cambridge University Press.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211–252.
- Colonus, H., & Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology*, 38(1), 35–58. 10.1006/jmps.1994.1002
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1), 2–5.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203.
- Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies (Doctoral dissertation). University of Connecticut.
- Frazier, L. (1987). Sentence processing: A tutorial review. In *Attention and performance XII: The psychology of reading* (pp. 559–586). London: Erlbaum.
- Frazier, L., & Clifton, C. J. (1997). Construal: Overview, motivation, and some new evidence. *Journal of Psycholinguistic Research*, 26(3), 277–295.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
- Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: Two sides of relative clause attachment. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 259–282). Cambridge, England: Cambridge University Press.
- Hesterberg, T., Moore, D. S., Clipson, A., & Epstein, R. (2005). Bootstrap methods and permutation tests. In D. S. Moore and G. P. McCabe (Eds.), *Introduction to the practice of statistics* (5th ed., pp. 1–70). New York: W.H. Freeman and Company.
- Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2002). Perspective effects on online text processing. *Discourse Processes*, 33(2), 159–173.
- Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, 4(1), 35–45.

- Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, 29(2), 241–248.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.
- Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, 51(5), 341–352.
- Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1231.
- McConkie, G. W., Rayner, K., & Wilson, S. J. (1973). Experimental manipulation of reading strategies. *Journal of Educational Psychology*, 66(1), 1–8.
- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29(2), 111–123.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283–312.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14(2), 247–279.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, 4(2), 61–64.
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1), 90–100.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308–313.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus*. New York: Springer.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raab, D. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24(5), 574–590.
- Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory and Cognition*, 36(1), 201–216.
- Townsend, J. T., & Ashby, G. F. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, England: Cambridge University Press.
- Townsend, J. T., & Colonius, H. (1997). Parallel processing response times and experimental determination of the stopping rule. *Journal of mathematical psychology*, 41(4), 392–397.
- Townsend, J. T., & Honey, C. J. (2007). Consequences of base time for redundant signals experiments. *Journal of Mathematical Psychology*(51), 1–24.
- Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39(4), 558–592.
- Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, 37, 513–525.
- van Gompel, R. P. G., & Pickering, M. J. (2006). Syntactic parsing. In M. J. Traxler, & M. A. Gernsbacher (Eds.), *The Oxford handbook of psycholinguistics* (2nd ed., pp. 289–307). London: Academic Press.
- van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process*. Oxford, England: Elsevier.
- 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 and Language*, 45(2), 225–258. 10.1006/jmla.2001.2773

- van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, 52(2), 284–307.
- Van Zandt, T. (2002). Analysis of response time distributions. *Psychonomic Bulletin and Review*, 1–99.
- Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82(4), 767.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S-PLUS*. New York: Springer.
- Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the SD of a response time distribution. *Psychological Review*, 114(3), 830–841.
- Wehrens, R., Putter, H., & Buydens, L. M. C. (2000). The bootstrap: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 54(1), 35–52.
- Wotschack, C. (2009). Eye movements in reading strategies (Doctoral dissertation). Universität Potsdam.