The processing domain of scope interaction

Oliver Bott & Fabian Schlotterbeck

Abstract

We present six experiments which investigated the time course of the interpretation of quantifier scope ambiguity. We used variable binding in the first quantifier to enforce scope inversion. The first two experiments were offline tests that established that the sentences had the proposed interpretations. The third experiment employed self-paced reading to show that potential difficulty in doubly quantified sentences with a variable is not due to interpreting the variable per se, but must come from scope inversion. We then report the results of two further self-paced reading experiments and an experiment in which we measured eye-movements during reading. Taken together, the reading time studies demonstrate that relative scope is only computed at the sentence-boundary. Only at the end of the sentence we found clear scope inversion effects. Our study shows that quantifier scope is not assigned incrementally, but depends on a complete minimal sentence.

1 Introduction

Sentence processing is largely incremental, that is, language comprehenders incorporate each word into the linguistic context as they encounter it (e.g., Marslen-Wilson (1973, 1975)). Evidence for incrementality comes from numerous studies showing that language comprehenders experience difficulty with temporary ambiguity occurring well before the end of the sentence. While the vast majority of the studies have investigated ambiguity at the syntactic or lexical level, sentential semantics hasn’t received that much attention. The present study addresses the incrementality of semantic interpretation by investigating the online resolution of quantifier scope ambiguities.

A sentence like (1-a) has two scope readings (1-b) versus (1-c), although its lexical items and surface structure are unambiguous. These can be paraphrased as follows. According to the inverse reading (1-b) there is a professor who has praised all his students, whereas according to the linear interpretation (1-c) for each student there is a (possibly different) professor having praised him.

(1)  a. Each student was praised by some professor.
    b. $\exists x[\text{professor}(x) \land \forall y[\text{student}(y) \rightarrow \text{praise}(x, y)]]$
    c. $\forall y[\text{student}(y) \rightarrow \exists x[\text{professor}(x) \land \text{praise}(x, y)]]$
As for processing, we may ask at which point during sentence comprehension relative scope is computed. In particular, are comprehenders able to determine scope even before they have encountered the verbal predicate, that is in the minimal configuration required for this type of ambiguity to show up. Immediate scope interpretation is captured by the following hypothesis.

**Incremental Scope Assignment (ISA):** When processing a multiply quantified sentence the processor immediately determines the relative scope of the quantifying expressions as they are encountered.

In opposition to the ISA, in formal semantics it is rather commonly assumed that the semantic representation is computed for the complete sentence lacking an incremental perspective\(^1\). For instance, in their classic textbook Heim and Kratzer (1998, 99 ff.) argue that the semantic representation is best computed top down, that is starting from the topmost node of the logical form of the sentence and working one’s way down the tree. Furthermore, quantificational determiners are usually taken to denote relations between sets – a determiner’s restriction and its nuclear scope (see eg., Peters and Westerståhl (2006)). In a simple quantificational sentence such as *every dog barks* the determiner *every* denotes a function that relates two sets: the set of dogs, ie. the quantifier’s restriction, and the set of barking entities, ie. its nuclear scope. According to this view, a determiner can be fully interpreted only when both restrictor and nuclear scope are available. Carrying this line of reasoning over to processing, quantifier interaction should only show effects after the quantifiers and the verbal predicate have been encountered. This is stated in the Global Interpretation Hypothesis.

**Global Interpretation (GIH):** When interpreting a quantifier the processor waits until both the restriction and the nuclear scope have been fully determined, that is the computation of scope will only start when the sentence is complete.

Although in most semantic accounts quantifier interaction is computed on the basis of a complete sentence, there are proposals which allow us to compute scope readings of not yet complete sentences, as well. One example is Barker (2003)’s continuation semantics. Consider the yet unfinished German sentence (2) up to *at least one pupil*.

(2) Jeder Lehrer hat mindestens einen Schüler ... (durchfallen lassen)
    Each teacher has at least one pupil ... (failed let)
    Each teacher has (failed) at least one pupil.

Leaving tense and grammatical aspect aside and thus ignoring the auxiliary, the first part of the sentence can be interpreted semantically using Barker’s rules for

---

\(^1\)These theories are usually not intended to make any predictions for processing. We will, however, rather naively take them as processing theories to see how far we get. Clearly, the derived predictions are our own and cannot be attributed to any of the theorists mentioned below.
linear or for inverse scope (cf. Barker, 2003, rules 18a/b). This yields a semantic representation in which quantifier scope is specified, but instead of the actual verbal predicate only a predicate variable is present. Thus, (2) reads either along the lines of every teacher is such that he did something to some pupil or some pupil is such that something was done to it by every teacher. We will not go into the formal details here, but hope this suffices as an illustration that there are semantic theories that allow us to model the computation of scope in an incremental fashion. This paper addresses the question whether such a theory is actually required.

The online resolution of quantifier scope ambiguity has been on the psycholinguistic agenda since Gillen (1991) and Kurtzman and MacDonald (1993). Most of the existing studies investigated the computation of quantifier scope using disambiguating continuation sentences with an anaphoric expression either in the singular (e.g., *this professor was very popular*) or in the plural (e.g., *these professors were very popular*) (Gillen, 1991; Kurtzman and MacDonald, 1993; Tunstall, 1998; Filik et al., 2004; Anderson, 2004; Zhou and Gao, 2009), but see Villalta (2003) for a different kind of disambiguation). The studies using this paradigm have shown that quantifier scope preference is influenced by various factors including structural, lexical, thematic and pragmatic properties of the quantifiers. One of them is the linear precedence of the quantifying expressions. If one quantifier precedes another, the former tends to take wide scope over the latter. Also, on the structural side, grammatical function has been found to influence scope preference with indirect objects taking wide scope more easily than direct objects (e.g. Filik et al., 2004). Yet another factor are the thematic properties of a quantifier. Since Ioup (1975) it has been repeatedly claimed that thematic relations constitute a scope hierarchy. The hierarchy predicts the tendency of agents to scope over patients; this has been supported by online experiments (e.g. Filik et al., 2004). Recently, Raffray and Pickering (2010) used priming to demonstrate that readers construct logical form (LF) representations during processing. The existence of priming between – with regard to lexical contents – unrelated sentences strongly supports the claim that LF representations are psychologically real.

To find out whether quantifier scope is computed incrementally it is crucial to measure effects of scope interaction well before the end of the sentence. Ideally, we would like to find a scope effect when the reader has just processed the quantifiers, but the rest of the sentence hasn’t been encountered yet. In the studies mentioned above effects were measured at the disambiguating region *this/these professor/s*, that is after comprehenders have already finished reading the scope ambiguous sentence. This is definitely too late to conclude anything about the incremental nature of scope interpretation.

To our knowledge, there are only three studies that have looked at effects well before the end of the sentence. One is the eyetracking study by Filik et al. (2004) also using singular/plural-disambiguations. They found differences in first-pass reading times of the ambiguous region, that is before the disambiguation. However, the second quantifier appeared in sentence final position and the critical region even crossed the clause boundary. This makes it impossi-
ble to tell what the processing domain (word, phrase, sentence) of quantifier scope really is. The second exception is a self paced reading study by Villalta (2003) investigating relative scope in how many questions (eg., how many sonatas did every pianist play at the concert?). The preceding context either forced a every > n – many or a n – many > every reading. Depending on the context, reading times differed. This effect showed up only at the end of the sentence even though, in principle, scope could have been computed already at every pianist, before play at the concert had been encountered. Finding a difference at such an early point would have been an indication of incremental interpretation². The lack of effect at the second quantifier region is, however, hard to interpret. Finally, in an eyetracking study on German inverse linking sentences like (3-a) vs. (3-b)³ Bott and Radó (2009) found longer first pass reading times at the second quantifier region on all pictures in (3-a) than at on each picture in (3-b) but no differences in non-quantificational controls like (3-c) vs. (3-d).

(3) a. Genau ein Tier auf allen Bildern sollst du nennen.
   Exactly one animal on all pictures should you name
   Name an animal on all pictures.

b. Genau ein Tier auf jedem Bild sollst du nennen.
   Exactly one animal on each picture should you name
   Name an animal on each picture.

c. Das Tier auf allen Bildern sollst du nennen.
   The animal on all pictures should you name
   Name the animal on all pictures.

d. Das Tier auf jedem Bild sollst du nennen.
   The animal on each picture should you name
   Name the animal on each picture.

Since the end of the sentence was highly predictable – it was always of the kind should you name – nothing can be concluded in terms of incrementality from this study, too. Moreover, it has been repeatedly claimed that the inverse linking construction is a scope island (May and Bale, 2005; Barker, 2003), and therefore isn’t a good test case in the first place. To sum up, none of the existing studies can inform us about whether scope is computed incrementally as predicted by the ISA.

The scope representations of yet incomplete sentences are particularly interesting from a processing perspective because they are highly abstract and cannot be easily connected to a concrete scenario. The ISA thus stands in opposition to what Sanford and Garrod (1998) put forward in their Scenario Mapping Fo-

² Although it would be unclear whether the preceding discourse context would have made the predicate predictable.

³ Instead of plain ein (a) they used genau ein (exactly one) which has the advantage that due to being non-monotonic the scope readings are logically independent of each other. Moreover, exactly one is clearly quantificational, whereas a or some are lexically ambiguous also displaying indefinite and specific interpretations. The present study therefore also used exactly one quantifiers.
Hypothesis: the interpretation of quantifier scope hinges on the concrete scenario comprehenders come up with when they encounter a particular verbal predicate and its arguments. According to Sanford and Garrod (1998) when, for instance, interpreting each room in the hotel has a bathroom comprehenders will immediately choose linear scope because the typical hotel does not have any rooms sharing bathrooms.

1.1 How Can We Test The Incremental Scope Assignment Hypothesis?

To investigate whether scope interpretation requires a complete sentence we need to compare constructions in which a quantifier phrase completes a(n) (at least) doubly quantified sentence with another construction in which it does not complete the sentence. A particularly interesting case is a construction where the verb has not been encountered yet. As already shown in (2), in a language like German verbs can appear sentence finally. We can thus test both $QP_1-V-QP_2$ and $QP_1-QP_2-V$ orders and measure integration difficulty of the second quantifier. Consider the sentences in (4) and (5).

   Each of his pupils acc praised exactly one teacher nom full-of goodwill.
   Exactly one teacher praised each of his pupils full of goodwill.

   Each of these pupils acc praised exactly one teacher nom full-of goodwill.
   Exactly one teacher praised each of these pupils full of goodwill.

c. Jeden seiner Schüler lobte dieser Lehrer voller Wohlwollen.
   Each of his pupils acc praised this teacher nom full-of goodwill.
   This teacher praised each of his pupils full of goodwill.

d. Jeden dieser Schüler lobte dieser Lehrer voller Wohlwollen.
   Each of these pupils acc praised this teacher nom full-of goodwill.
   This teacher praised each of these pupils full of goodwill.

In (4-a) and (4-b) the topicalized object quantifier precedes the subject quantifier genau ein Lehrer (exactly one teacher). With these quantifiers, linear scope is clearly preferred Pafel (2005). However, in (4-a) inverse scope is required in order to interpret seine (of his) as a bound variable. In (4-b) having dieser (of these) instead of seiner no variable binding is required and readers can maintain the preferred linear reading. By contrast, integrating dieser Lehrer (this
teacher is as easy in (4-c) as it is in (4-d) because it is a referential expression taking widest scope (e.g., Fodor and Sag, 1982; Beghelli and Stowell, 1997). Thus, interpreting the possessive shouldn’t cause any problem in these examples.

(5)  
   Each of his pupils has exactly one teacher full of goodwill.  
   Exactly one teacher has praised each of his pupils full of goodwill.  
b. Jeden dieser Schüler hat genau ein Lehrer voller Wohlwollen gelobt.  
   Each of these pupils has exactly one teacher full of goodwill.  
   Exactly one teacher has praised each of these pupils full of goodwill.  
c. Jeden seiner Schüler hat der Lehrer voller Wohlwollen gelobt.  
   Each of his pupils has the teacher full of goodwill.  
   The teacher has praised each of his pupils full of goodwill.  
d. Jeden dieser Schüler hat der Lehrer voller Wohlwollen gelobt.  
   Each of these pupils has the teacher full of goodwill.  
   The teacher has praised each of these pupils full of goodwill.

The sentences in (5) are exactly like those in (4) except for one crucial difference. In (5), the verb praise appears only at the end of the sentence. Nevertheless, the scope of the quantifiers in (5-a) versus (5-b) can, in principle, be determined before having accessed the verbal information. Variable binding in (5-a) forces inverse scope, whereas the anaphor dieser (of these) in (5-b) doesn’t. If scope interpretation proceeds strictly incrementally in line with the ISA, we expect difficulty at the second quantifier phrase regardless of the position of the verb. If, on the other hand, scope is assigned globally, we expect no effects due to variable binding at the second quantifier region, but only delayed effects when the verbal predicate is eventually encountered.

Below, we report two self-paced reading and an eyetracking experiment testing the ISA hypothesis in sentences like (4) and (5). Before coming to these experiments (Experiments 4a/b and 5) we will reflect upon some assumptions that were tacitly made when we introduced the design of our study. We will show that linear scope is indeed preferred in German doubly quantified object before subject sentences of the type (4-b)/(5-b). Secondly, we will provide evidence that sentences like (4-a) and (5-a) receive overwhelmingly bound variable interpretations. Finally, we will see that potential difficulty cannot be attributed to a bound variable interpretation per se, but must be due to the computation of inverse scope.
1.2 Underlying Assumptions

Although there is an intuitive preference for linear scope in German sentences like (4-b) and (5-b) (see Pafel, 2005, for a comprehensive work on quantifier scope in German), this intuitive preference has to be further validated by experimental results. In particular, we have to choose a construction in which linear scope is preferred, but inverse scope is still possible. For this purpose we tested two constructions with topicalized objects in a picture verification task and selected the one that best fits these requirements. The first experiment (Experiment 1) tested the following assumption (A1): when interpreting doubly quantified sentences like (4-b) with an object quantifier preceding the subject quantifier linear scope is preferred.

Secondly, when introducing (4-a) and (5-a) we tacitly assumed that a bound variable interpretation of *seiner* (*his*) is the only possible interpretation of the possessive pronoun. But this is clearly not the case. Consider the following discourse under a coreferential interpretation as indicated by the indices.

(6) [Professor Brown] is an exceptional teacher. Each of his students was praised by some colleague.

The example shows that with the right context, *his* can receive a coreferential interpretation. What, if readers are able to come up with the right kind of context even in isolated sentences like (4-a) and (5-a)? This seems particularly likely given the results of a study by Frazier and Clifton (2000). Testing English sentences of the type *According to Ann, everybody loves his sister* they found that readers preferably interpreted *his sister* with an extra-sentence antecedent. To check whether this is the case in our sentences, too, we have to show that the possessive pronoun clearly disambiguates conditions (4-a) and (5-a) towards an inverse interpretation and that comprehenders do not try to resolve the pronoun sentence externally. The following assumption (A2) was tested in Experiment 2: when presented out of context, comprehenders interpret sentences (4-a) and (5-a) with a bound variable interpretation.

Finally, we have to make sure that enhanced difficulty in (4-a) and (5-a) is due to assigning inverse scope and not to general difficulty when readers have to interpret a bound variable. This assumption (A3) was tested in Experiment 3. It is important, because we need to be sure that the presumed difficulty of (4-a)/(5-a) is due to computing a dispreferred scope interpretation.

2 Experiment 1: Linear Scope Is Preferred Over Inverse Scope

We need a doubly quantified construction in which linear scope is preferred. Why should the inverse interpretation in object topicalized (OVS) sentences like (4-b) be dispreferred? This is a crucial question because some theories of quantifier scope (for instance Frey, 1993) predict German object topicalized sentences like (7) to be fully ambiguous.
Almost every pupil$_{\text{acc}}$ praised at least one teacher$_{\text{nom}}$. At least one teacher praised almost every pupil.

The available readings reported in the literature are, however, sometimes quite controversial. For instance, (8) has been claimed both to be fully ambiguous (May, 1977; Hornstein, 1984; Higginbotham, 1985), and to only allow the $\forall \exists$-reading (Reinhart, 1976, 1983; Hornstein, 1995; Beghelli and Stowell, 1997).

(8) Everyone loves someone.

As for (7), the available readings aren’t clear either. Reinhart (1983)’s account, for instance, is more restrictive than Frey (1993)’s and, due to c-command relations in s-structure, predicts the sentence to be unambiguous with linear scope, even though in later work (Reinhart, 1995) weakens the c-command condition so that it only supplies the unmarked or most accessible reading. The latter position is supported by psycholinguistic investigations on the online processing of quantifier scope (e.g. Tunstall, 1998; Anderson, 2004) providing evidence that the preferred interpretation of a doubly quantified sentence corresponds to the logical form that minimally differs from surface structure.

To resolve some of these issues, the first experiment tested which scope interpretation (linear or inverse) is preferred in German doubly quantified sentences of two different kinds, and whether the inverse reading is only dispreferred or completely unavailable. We tested two types of object topicalized doubly quantified German sentences like in (9-a) and (9-b) (the former repeated from (4-b)).

(9) a. Jeden dieser Schüler lobte genau ein Lehrer voller
   [Each of these pupils]$_{\text{acc}}$ praised [exactly one teacher]$_{\text{nom}}$ full of goodwill.
   Exactly one teacher praised each of these pupils full of goodwill.

b. Genau einen Schüler lobte jeder Lehrer voller
   [Exactly one pupil]$_{\text{acc}}$ praised [each teacher]$_{\text{nom}}$ full of goodwill.
   Each teacher praised exactly one pupil full of goodwill.

The semantic literature on quantifier scope does not provide us with uncontroversial predictions for the two constructions. Theories like Reinhart (1983) or Frey (1993) take the syntactic configuration to be the sole factor that is relevant for quantifier scope. According to these theories, and many others, both (9-a) and (9-b) are expected to be fully ambiguous. Other theories, for instance Beghelli and Stowell (1997) or Pafel (2005) predict semantic properties like distributivity or discourse linking of quantifiers to also influence their scope taking behavior. According to these theories, the preference for linear scope should be stronger in (9-a) than in (9-b), because distributive each of these can outscope exactly one more easily than vice versa. In particular Pafel (2005)’s
multifactorial scope theory allows us to derive very strong predictions for these examples, namely that (9-b) should display a slight preference for an inverse interpretation, whereas (9-a) should be interpretable exclusively on its surface scope interpretation.

How can we measure scope preferences and tell a dispreferred interpretation from an unavailable one? We employed a picture verification task which is, just like its close relative, the truth-value judgment task, well suited to examine the exact status of dispreferred readings (cf. Crain and Thornton, 1998). As for disambiguation, we drew upon the work by Bott and Radó (2007) who compared three different offline tasks to measure scope preference. In their study, a picture verification task using set diagrams turned out to be the most reliable and valid mode of disambiguation among the three methods tested in a cross-methodological comparison.

In order to have clear baseline controls, we accompanied the constructions under consideration (9-a)/(9-b) with clearly disambiguated sentences in which the quantifiers were separated by a clause boundary. This provided us with a criterion of when to consider a sentence scope unambiguous: the potentially ambiguous construction is in fact unambiguous if it is indistinguishable from a scope disambiguated control.

2.1 Methods

2.1.1 Materials

We constructed a set of 36 doubly quantified sentences in the two sentence conditions each. (10) is the sample item repeated from (9-a)/(9-b).

(10) a. Genau einen Schüler lobte jeder Lehrer voller Wohlwollen.
   Exactly one pupil praised each teacher full of goodwill.
   Each teacher praised exactly one pupil full of goodwill

b. Jeden dieser Schüler lobte genau ein Lehrer...
   Each of these pupils praised exactly one teacher...
   Exactly one teacher praised each of these pupils...

We added two scope unambiguous controls which were created by separating the quantifiers each (∀) and exactly one (∃!) by a clause boundary. (11) illustrates the control conditions.

   Exactly one pupil holds: him praised each teacher full of goodwill
   Exactly one pupil is such that he was praised by each teacher full of goodwill.

4The experiment was part of a larger experiment with a total of 12 conditions also testing canonical subject before object sentences. Since these aren’t relevant for our current purposes we will only focus on the eight object before subject conditions.
Each of the four sentence conditions was paired with two pictures which were either compatible with the $\forall \exists$ or the $\exists \forall$ reading, but not vice versa. This yielded a total of eight conditions in a $2 (construction: \text{exactly one–each}}$ vs. $\text{each–exactly one}) \times 2 (ambiguity: \text{ambiguous vs. disambiguated}) \times 2 (diagram: \text{linear vs. inverse})$ within design. Sample pictures for both constellations of quantifiers are illustrated in Figures 1(a)–1(c). The $\forall \exists$-picture was the same in both sentence types. The $\exists \forall$ pictures in the $\text{exactly one}$ before $\text{each}$ order (10-a) and (11-a) were just the mirror image of those used in the $\text{each of these}$ before $\text{exactly one}$ order (10-b) and (11-b). We added 61 filler sentence-picture pairs (15 false) and distributed the materials to lists according to a latin square design.

### 2.1.2 Procedure and Participants

The experiment was implemented in WebExp 2 (Mayo et al., 2006a). After reading written instructions, participants first had to do a short practice session of ten trials. Then followed the experiment in a single block. Sentence-picture pairs were presented in individually randomized order without time pressure. Participants were individually tested in a quiet computer pool at the department of modern languages.

Judgments had to be provided on a seven point scale. A value of 1 indicated that the sentence was inconsistent with the picture, whereas a value of 7 indicated that it was a perfectly natural description of the picture. The ratings were normalized by computing z-scores for each participant.

48 students from Tübingen University (mean age 24.7y.; range 20–33y.; 32 female) took part in the study for a payment of 5€. An experimental session took about 30 minutes.
2.2 Results and Discussion

Mean judgments are depicted in Figure 2. We will report separate statistical analyses for the two constructions. Each was analyzed by computing 2 (ambiguity) x 2 (disambiguating picture) repeated measures ANOVAs.

The first construction type, in line with Pafel (2005)'s predictions, was compatible with both scope readings and the inverse reading was somewhat preferred over the linear interpretation (mean z-scores: 0.19 vs. -0.09). The scope unambiguous control conditions patterned with the true and false fillers, respectively. The scope disambiguated sentences were judged true when combined with a picture only compatible with the linear interpretation and judged false when combined with an inverse picture. ANOVAs revealed significant main effects of ambiguity (\( F_1(1, 47) = 35.0, p < .01 \); \( F_2(1, 35) = 52.1, p < .01 \)) and of diagram (\( F_1(1, 47) = 74.2, p < .01 \); \( F_2(1, 35) = 98.9, p < .01 \)) and a significant interaction between ambiguity and diagram (\( F_1(1, 47) = 109.4, p < .01 \); \( F_2(1, 35) = 197.2, p < .01 \)). We computed a pairwise comparison between the linear and the inverse ambiguous conditions to test whether the preference for the inverse interpretation was statistically reliable. Paired t-tests revealed that the difference was marginal by participants and significant by items (\( t(47) = 1.88, p = .07 \); \( t(35) = 2.58, p < .05 \)). Thus, this type of construction is not suited for the aims of the present paper because it is not clear whether it will cause measurable processing difficulty if disambiguated towards the inverse reading.

The second type of doubly quantified sentences, however, meets the requirements of the first assumption (A1). In the ambiguous construction linear scope was strongly preferred over inverse scope (mean z-scores: 0.44 vs.
-0.58) but the difference between the two readings was less pronounced than in the disambiguated sentences (mean z-scores: 0.57 (linear) vs. -0.82 (inverse)). Thus, although highly dispreferred the inverse reading is still possible in this type of construction. ANOVAs revealed a significant main effect of \textit{diagram} ($F_1(1, 47) = 165.57, p < .01$; $F_2(1, 35) = 345.19, p < .01$) and a significant interaction between \textit{ambiguity} and \textit{diagram} ($F_1(1, 47) = 10.24, p < .01$; $F_2(1, 35) = 8.15, p < .01$) but the main effect of \textit{ambiguity} was not significant ($F_{1/2} < 1$). The interaction is due to the fact that the difference between the judgments for the two scope readings was bigger in the unambiguous cases than in the ambiguous ones.

This experiment aimed at establishing that German doubly quantified OVS sentences of the type discussed in the introduction are scope ambiguous with a clear preference for surface scope. The findings clearly provide evidence for assumption A1. In all of the experiments to follow we therefore used the constructions with each of these preceding exactly one.

3 Experiment 2: Bound Variable Readings

What happens to the interpretation of a doubly quantified OVS sentence such as (12-a) (repeated from (4-b)) when the first quantifier contains a possessive pronoun as in (12-b) (repeated from (4-a))? 

   Each of these pupils\textsubscript{acc} praised exactly one teacher\textsubscript{nom}.
   Exactly one teacher praised each of these pupils.

   Each of his pupils\textsubscript{acc} praised exactly one teacher\textsubscript{nom}.
   Exactly one teacher praised each of his pupils.

In (12-b) the possessive pronoun either needs a binder or an anaphoric link to some referent from preceding discourse (henceforth extra-sentence coreferential interpretation). In the introduction we simply assumed that a bound variable interpretation is preferred over the accommodation of a yet missing discourse referent. As it turns out, this assumption is not uncontroversial.

The hypothesis that bound variable interpretations are preferred over coreferential interpretations was supported for children by Avrutin (1994, 1999). In particular, he showed that children can master syntactic conditions on binding even before they learned to apply discourse constraints (see also Guo et al. (1996) and Vasić (2006) for similar results also in agrammatic aphasics). Further empirical support for the preference of bound variable interpretations over coreferential interpretations comes from the processing of VP ellipsis in adults. In a self-paced reading study on the interpretation of bound variables in English Frazier and Clifton (2000, Exp. 1a/b) found that adult readers spent longer on sentences that required a coreferential interpretation (John thinks it’s a good idea to shave himself before he goes to bed and Anne does too) than on sentences also allowing a bound variable interpretation (John thinks it’s a
good idea to shave himself before he goes to bed and Andy does too). Extending their research to construction types other than VP ellipsis they, however, found rather surprising results (see Burkhardt (2005) for similar findings regarding “referential quantifiers”). Most relevant for our purposes are sentences like the one in (13).

(13) According to Ann, everyone loves his sister.

Frazier and Clifton (2000, Exp. 2a) let participants choose paraphrases that either disambiguated towards a bound variable reading (his = everyone’s), an intra-sentence coreferential reading (his = Ann’s) or an extra-sentence coreferential reading (his = somebody else’s). Difficulty ratings showed that sentences with gender information blocking intra-sentence coreferential interpretations were harder to understand than sentences that did not block them (eg. According to Sam...). Moreover, for sentences like (13) extra-sentence paraphrases were even preferred over bound variable paraphrases. Thus, even when presented completely out of context sentences with a potential binder and a possessive pronoun seem to very often get accommodated and receive an extra-sentence coreferential interpretation instead of a bound reading. A possible concern is, however, that the paraphrase for the extra-sentence coreferential reading (eg. according to Ann, there is some guy (say Bill) and everyone loves Bill’s sister) might have introduced a discourse referent triggering the extra-sentence coreferential interpretation in the first place.

In contrast to Frazier and Clifton (2000)’s findings, Koornneef and colleagues (Koornneef et al., 2006; Koornneef, 2008, 2010) showed in a number of eyetracking studies that bound interpretations are easier to process than coreferential interpretations. They tested constructions like (14) in Dutch by putting them in discourse contexts that either biased towards a bound or a coreferential interpretation. The reading times clearly indicated that these sentences were harder to process in the coreferential than in the bound variable contexts.

(14) Every worker who knew that Paul was running out of energy, thought it was very nice that he could go home early this afternoon.

None of the existing studies has investigated bound vs. coreferential interpretations in German. We therefore conducted an experiment to measure binding preferences in the constructions under study. It employed a paraphrase selection task in which we provided paraphrases for each of the theoretically possible readings of doubly quantified sentences with (12-b) and without a possessive pronoun (12-a) in the first quantifier.

Another question that was addressed in the present experiment was whether comprehenders would employ a wait-and-see strategy when they have to deal with sentences like (12-b). It seems reasonable that the processor will wait for a discourse referent that is yet to come instead of reversing scope in order to establish binding. Following this line of reasoning, bound variable readings would only be computed at a point when there is clear evidence that no more referents will be introduced, that is at the end of a discourse segment. Obviously,
in reading times this would look like a late effect, even though it would tell us nothing about the online interpretation of quantifier scope. To find out about the particular strategies comprehenders have, we therefore embedded the constructions of interest in sentences that introduced an additional discourse referent. Consider (15).

(15) Each of his students was greeted by some professor, said Peter.

Here, his can be either interpreted as bound by exactly one teacher or it can be understood as coreferring to Peter. If comprehenders avoid scope inversion due to its inherent processing costs, but instead wait for a coreferent that is yet to come, we would expect to find no bound interpretations in constructions like these. If we find a considerable proportion of bound variable interpretations, however, we can be sure that comprehenders notice the possibility of binding under inverse scope.

3.1 Methods

48 native German speakers (mean age 24.4 y., range 20-51 y., 32 female) from Tübingen University took part in the experiment for monetary compensation of 8 €. It was a web-administered paraphrase selection task implemented in WebExp 2 (Mayo et al., 2006b). Participants were tested individually without any time pressure. In each trial, they had to read a sentence and choose one out of three paraphrases which were displayed right beneath the sentence. If they felt that neither of these fit they could choose “neither fits”, instead. The order in which the paraphrases appeared in the display was counterbalanced so that in each condition each type of paraphrase appeared equally often in each position. After written instructions participants did a practice session of eight trials, then the experiment followed in a single block. The sentence materials appeared in individually randomized order. An experimental session lasted approximately 45 minutes.

3.1.1 Materials

We modified the first 32 ambiguous OVS sentences from Experiment 1 and created eight conditions. The sample item in (16) illustrates the non-embedded conditions. We manipulated the within factors pronoun (with possessive vs. without possessive) and position of the full verb (v-second vs. v-end). The latter manipulation was included to test whether the position of the verb affects binding and scope.

    Each of his pupilsacc praised exactly one teachernom.
    Exactly one teacher praised each of his pupils.

    Each of these pupilsacc praised exactly one teachernom.
    Exactly one teacher praised each of these pupils.
c. Jeden seiner Schüler hat genau ein Lehrer gelobt.
   Each of his pupils has exactly one teacher praised.
   Exactly one teacher praised each of his pupils.

d. Jeden dieser Schüler hat genau ein Lehrer gelobt.
   Each of these pupils has exactly one teacher praised.
   Exactly one teacher praised each of these pupils.

The sentences in (17) illustrate the four embedded conditions. The doubly quantified sentences with a possessive appeared in sentences that either had a potential coreferent (cf. (17-a)/(17-c)) or had a referent that was of the wrong gender (cf. (17-b)/(17-d)). The latter conditions were included as a baseline.

We also manipulated the position of the full verb (v-second vs. v-end).

   Each of his pupils praised exactly one teacher, told Hans.
   Hans told that exactly one teacher praised each of his pupils.

b. Jeden seiner Schüler lobte genau eine Lehrerin, sagte Hans.
   Each of his pupils praised exactly one female teacher, told Hans.
   Hans told that exactly one female teacher praised each of his pupils.

   Each of his pupils has exactly one teacher praised, told Hans.
   Hans told that exactly one teacher praised each of his pupils.

d. Jeden seiner Schüler hat genau eine Lehrerin gelobt,
   Each of his pupils has exactly one female teacher praised, told Hans.
   Hans told that exactly one female teacher praised each of his pupils.

Each item in each condition was paired with three paraphrases. To simplify matters, we only report paraphrases for the conditions with the main verb in second position. The verb-final paraphrases were formulated analogously.

The three theoretically possible readings of the non-embedded conditions (16-a) and (16-c) are illustrated in (18-a)-(18-c). (18-a) paraphrases the bound, inverse interpretation, (18-b) is a coreferential, sentence-external interpretation with linear scope, while (18-c) is a coreferential, sentence-external interpretation with inverse scope.

(18) a. Exactly one teacher praised each of the own pupils.

b. There is a certain teacher, say Mr. Lämpel, and each of Lämpel’s pupils was praised by exactly one teacher.

c. There is a certain teacher, say Mr. Lämpel, and exactly one other teacher praised each of Lämpel’s pupils.
Paraphrases for the non-embedded *qq-this* conditions (16-b) and (16-d) are exemplified in (19-a)-(19-c). Besides the linear scope paraphrase in (19-a) and the inverse scope paraphrase in (19-b) we included a dummy paraphrase (19-c) to keep the number of paraphrases identical across conditions. The dummy paraphrases were generated by interchanging the quantificational determiners *each of these* and *exactly one* (cf. (19-a) vs. (19-c)).

\[(19)\]  
\begin{align*}
& a. \text{ Each of these pupils is such that he was praised by exactly one} \\
& \text{ teacher.} \\
& b. \text{ Exactly one teacher is such that he praised each of these pupils.} \\
& c. \text{ Each of these teachers is such that he praised exactly one pupil.}
\end{align*}

The embedded *qq-his* conditions (17-a) and (17-c) received a paraphrase for the bound interpretation (20-a) and two coreferential paraphrases. The first (20-b) paraphrases a linear coreferential interpretation whereas the second is an inverse coreferential paraphrase (20-c). The paraphrases for the embedded *qq-his* baseline conditions (17-b) and (17-d) were identical to the ones in (20-a)-(20-c). In this case, however, the paraphrase in (20-a) is a dummy paraphrase which should provide us with a baseline control. Note that due to a gender mismatch between *seiner (his)* and *Lehrerin (female teacher)* the possibility of a bound interpretation is ruled out in this condition.

\[(20)\]  
\begin{align*}
& a. \text{ Hans said that exactly one teacher praised each of the own pupils.} \\
& b. \text{ Hans said that each of Hans' pupils was praised by exactly one} \\
& \text{ teacher.} \\
& c. \text{ Hans said that exactly one teacher praised each of Hans' pupils.}
\end{align*}

We added 48 fillers which superficially resembled the experimental items together with their paraphrases and constructed eight lists in a latin square design.

### 3.1.2 Data Analysis

The non-embedded and the embedded conditions were analyzed separately from each other. The non-embedded conditions were analyzed with respect to the factors *paraphrase* (*bound/dummy* vs. *linear* (*coreferential*) vs. *inverse* (*coreferential*) vs. *none fits*), *pronoun* (*with possessive* vs. *without possessive*) and *verb position* (*v-second* vs. *v-end*). In the embedded conditions we were only interested in the availability of bound variable interpretations in the +binding conditions. We therefore analyzed the proportions of bound vs non-bound interpretations, that is the sum of the three non-bound cases, and dummy coded them as a new factor *bound interpretation*. The embedded conditions were analyzed with respect to *bound interpretation* (*bound/dummy* vs. *other paraphrases*), *binding* (*binding vs. no binding*) and *verb position* (*v-second* vs. *v-end*). Inferential analyses were based on hierarchical log-linear models (see Scheepers (2003), for discussion). Participants and items were entered, separately, as factors in the computation of partial association Likelihood Ratio Chi-Squares (*LRC*; in the
Table 1: Distribution of selected paraphrases in the non-embedded conditions of Exp. 2. In the conditions without a possessive pronoun the first type of paraphrase were dummy paraphrases. In the conditions with a pronoun the first type were bound variable paraphrases and the linear and inverse paraphrases corresponded to (sentence external) coreferential linear and inverse scope readings. “Other” are the proportion of “neither fits” button presses.

<table>
<thead>
<tr>
<th>Pronoun</th>
<th>Verb Position</th>
<th>Paragraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Possessive</td>
<td>Q-V-Q</td>
<td>13.9% 55.6% 21.5% 9.0%</td>
</tr>
<tr>
<td></td>
<td>Q-Aux-Q-V</td>
<td>8.3% 61.1% 25.0% 5.6%</td>
</tr>
<tr>
<td>+ Possessive</td>
<td>Q-V-Q</td>
<td>78.5% 6.3% 9.0% 6.3%</td>
</tr>
<tr>
<td></td>
<td>Q-Aux-Q-V</td>
<td>77.1% 8.3% 5.6% 9.0%</td>
</tr>
</tbody>
</table>

following LRCS\textsubscript{1} refers to the analysis by participants and LRCS\textsubscript{2} to the analysis by items. This enables the generalizability of main effects and interactions to be assessed across items and participants. Hierarchical loglinear models are particularly suited to the analysis of situations in which frequency data in one cell (eg. bound interpretations) are contingent on frequency data in other cells (eg. linear or inverse coreferential interpretations).

3.2 Results

3.2.1 Non-embedded Sentences

Table 1 presents the distribution of paraphrases selected in the four non-embedded conditions. In the two conditions with a possessive pronoun participants overwhelmingly (77.8% of the time) chose bound variable paraphrases with inverse scope. In less than 15% of all cases they selected one of the coreferential interpretations which were, in principle, valid interpretations of the sentences. In the conditions without a possessive pronoun they overwhelmingly (81.6% of the time) chose either linear or inverse scope paraphrases, but linear scope was preferred over inverse scope (60.9% vs. 23.25%). The proportions of selected paraphrases were similar across the two word orders. Verb position thus didn’t seem to have an effect.

This was corroborated by the statistical analyses. The two-way interaction of pronoun and paraphrase was highly significant (LRCS\textsubscript{1} = 364.3, df = 3, p < .01; LRCS\textsubscript{2} = 332.9, df = 3, p < .01). This interaction is due to the different distribution of selected paraphrases in the conditions without a possessive pronoun compared to the conditions with a possessive pronoun. Neither the three-way interaction of pronoun, verb position and paraphrase (LRCS\textsubscript{1} = 5.82, df = 3, p = .12; LRCS\textsubscript{2} = 5.28, df = 3, p = .15) or the two-way interaction of verb position and paraphrase (LRCS\textsubscript{1} = 2.22, df = 3, p = .53; LRCS\textsubscript{2} = 1.02, df = 3, p = .80) were reliable. Thus, the distribution of selected paraphrases was the same across word orders.
3.2.2 Embedded Sentences

Table 2 presents the distribution of paraphrases selected in the four embedded conditions. In the +binding conditions the bound variable interpretation was chosen 32% of the time. Compared to the –binding control conditions this corresponds to an 21.4% increase in bound interpretations. In the statistical analysis this increase resulted in a significant interaction of binding and bound interpretation ($LRCS_1 = 49.99$, $df = 1$, $p < .01$; $LRCS_2 = 54.16$, $df = 1$, $p < .01$). Verb position again had no effect; neither the two-way interaction of verb position and bound interpretation ($LRCS_1 = .01$, $df = 1$, $p = .95$; $LRCS_2 = .01$, $df = 1$, $p = .94$) or the three-way interaction of verb position, binding and bound interpretation ($LRCS_1 = 1.16$, $df = 1$, $p = .28$; $LRCS_2 = .07$, $df = 1$, $p = .79$) was significant.

3.3 Discussion

The present experiment provides offline evidence that the construction to be tested in the following reading time experiments meets assumption A2. First of all, in the non-embedded conditions without a possessive pronoun we observed a clear preference for linear scope, but inverse scope was still available as indicated by more than 20% inverse scope paraphrases. The presence of a possessive pronoun disambiguated the sentences towards a bound variable interpretation. The coreferential alternatives were almost never chosen, even though the coreferential paraphrases introduced an additional referent which should facilitate a coreferential extra-sentence interpretation. This lets us expect that comprehenders will compute a bound interpretation for the doubly quantified constructions with the possessive if they appear without context. Embedding these constructions in a sentence that made a referent available for coreference revealed that a bound interpretation is not just a last resort, but a viable option that is considered even under competition with a coreferential alternative.

Above that, our results indicate that manipulating the position of the main verb does not affect the meanings of the sentences under consideration. This is an important prerequisite for the reading time experiments. Interpretation pref-

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>– Binding</td>
<td>Q-V-Q</td>
<td>11.8%</td>
<td>36.8%</td>
<td>35.4%</td>
<td>16.0%</td>
</tr>
<tr>
<td></td>
<td>Q-Aux-Q-V</td>
<td>9.7%</td>
<td>41.0%</td>
<td>33.3%</td>
<td>16.0%</td>
</tr>
<tr>
<td>+ Binding</td>
<td>Q-V-Q</td>
<td>30.6%</td>
<td>30.6%</td>
<td>27.8%</td>
<td>11.1%</td>
</tr>
<tr>
<td></td>
<td>Q-Aux-Q-V</td>
<td>33.3%</td>
<td>26.3%</td>
<td>27.1%</td>
<td>13.2%</td>
</tr>
</tbody>
</table>

Table 2: Distribution of paraphrases in the embedded conditions of Exp. 2. In the –binding conditions the first type of paraphrase was a dummy paraphrase, whereas in the +binding conditions the same paraphrase characterized a bound variable interpretation. Coref. Lin. = coreferential linear; Coref. Inv. = coreferential inverse.
erences are the same, but the amount of information when the second quantifier
is encountered differs drastically between the two word order variants.

Our findings nicely fit the results of Koornneef and colleagues. Like them
we found a clear preference for bound variable interpretations over coreferential
interpretations. Dutch and German thus seem to behave rather similar in this
respect. At first sight, our results, however, seem to be incompatible with
Frazier and Clifton (2000)’s English binding data. Reconsidering their sentence
materials repeated in (21) reveals a crucial difference between the studies.

(21) According to Ann, everyone loves his sister.

In order to achieve a bound variable interpretation of (21), everyone has to be restricted to everyone who has a sister. This restriction comes as a presupposition that has to be accommodated. It is a rather natural assumption that accommodation is a costly operation (eg. Altmann and Steedman (1988)) and therefore comprehenders may have shifted to a coreferential reading which might be easier under these special circumstances. It has to be seen whether this really is the case. Unfortunately, we didn’t have access to the full set of materials to check whether all their sentences had this property.

To sum up, the first two experiments provide offline evidence that the chosen constructions in fact serve their purpose. Experiment 1 confirmed that doubly quantified German OVS sentences are indeed scope ambiguous, albeit in the type of construction to be used in the online experiments the linear interpretation is strongly preferred over inverse scope. Thus, we should expect to find a scope-inversion effect when the processor integrates the second quantifier. Experiment 2 showed that the possessive pronoun in the first quantifier receives a bound variable interpretation which is only possible under inverse scope. Thus, our first two assumptions are borne out. Before testing the ISH, however, we have to establish the third assumption: in the constructions under study, a bound variable interpretation is not hard per se, but potential difficulty must be due to a shift in quantifier scope. This assumption was tested in the next experiment.

4 Experiment 3: Bound Variables Are Not Necessarily Difficult

In (22-a) and (22-b) the quantifiers appear in the opposite order than in the constructions (4) and (5) that will be used to investigate the ISH.

(22) a. Genau ein Lehrer lobte jeden seiner Schüler...
   Exactly one teacher\textsubscript{nom} praised each of his pupils\textsubscript{acc}...
   Exactly one teacher praised each of his pupils...

b. Genau ein Lehrer lobte jeden dieser Schüler...
   Exactly one teacher\textsubscript{nom} praised each of these pupils\textsubscript{acc}...
   Exactly one teacher praised each of these pupils...
According to our third assumption, interpreting the bound variable *his* in (22-a) shouldn’t pose any problems because in the linear interpretation the binding quantifier *exactly one teacher* scopes over *each of his pupils* and binding does not require inversion. This can be controlled for if we compare the processing of (22-a) with (22-b) which doesn’t require any binding. Intuitively, (22-a) does not feel any harder than the control (22-b), but if we take a closer look at online studies of pronoun interpretation it becomes clear that there may be differences in processing complexity between bound variables and anaphoric expressions.

Evidence for bound variables to be more complex than coreferential pronouns comes for instance from Carminati et al. (2002), who conducted two eyetracking experiments comparing reading times of pronouns with a quantificational antecedent (eg. *every British soldier*) to pronouns with coreferential antecedents (eg. *the old British soldier*) and to pronouns with a proper name antecedent. They found increased rereading time for the quantified pronoun compared to the coreferential one. At first glance, this suggests that bound pronouns are harder to comprehend than coreferential pronouns. However, the sentences also differed with respect to whether they were quantificational or not. Thus, the difference in reading time could also be due to quantified sentences being harder to comprehend than sentences with a definite description.

In a cross modal lexical decision task Burkhardt (2005) also found higher RTs for bound pronouns in Dutch quantificational sentences than for coreferential pronouns at a probe position after readers had encountered the pronoun, whereas at an earlier probe position before the pronoun there was no difference. This lends additional support to the claim that bound pronouns are more difficult to comprehend than coreferential pronouns. However, the sentences also differed with respect to whether they were quantificational or not. Thus, the difference in reading time could also be due to quantified sentences being harder to comprehend than sentences with a definite description.

Evidence in this direction comes from the experiments of Koornneef (2008). In one of his experiments (Exp. 5), he showed that bound pronouns in disambiguated Dutch versions of discourses like (24-a) were as easy to comprehend as the accompanying coreferential pronoun condition in (24-b). In general, his experiments established a preference of bound variable interpretations over coreferential interpretations.

(24)  a. A war was going on in Sudan and the soldiers at the front were

Another factor that might have made (23-a) more complex than (23-b) is an issue that was already raised in the discussion of the last experiment. The quantificational sentence requires restriction which is not explicitly provided in the sentence. The reader has to infer *everyone* to be *everyone who has clients* to make sense of it. This isn’t the case in (23-b).

Evidence in this direction comes from the experiments of Koornneef (2008). In one of his experiments (Exp. 5), he showed that bound pronouns in disambiguated Dutch versions of discourses like (24-a) were as easy to comprehend as the accompanying coreferential pronoun condition in (24-b). In general, his experiments established a preference of bound variable interpretations over coreferential interpretations.
constantly thinking about death. Every soldier was afraid that he was going to die on the bloody battlefield.

b. A war was going on in Sudan and a soldier at the front was constantly thinking about death. The soldier was afraid that he was going to die on the bloody battlefield.

To sum up, although some of the existing studies on bound pronouns lend support to the claim that bound pronouns might be harder to comprehend than coreferential expressions, the evidence isn’t entirely clear and calls for additional experimental work which separates quantificational costs from the cost of binding. This was the aim of the present experiment.

4.1 Methods

4.1.1 Materials

The 32 items from the previous experiment were modified by changing the order of quantifiers to SVO. They were segmented as in (25-a) and (25-b).

   Exactly one teacher | praised | each of-his pupils | full-of | goodwill.
   Exactly one teacher praised each of his pupils full of goodwill.

b. Genau ein Lehrer | lobte | jeden dieser Schüler | voller | Wohlwollen.
   Exactly one teacher | praised | each of-these pupils | full-of | goodwill.
   Exactly one teacher praised each of these pupils full of goodwill.

Both conditions (25-a) and (25-b) contained a pronoun. Whereas his is bound by the first quantifier, the demonstrative pronoun these calls for an accommodation of a set of salient pupils. This shouldn’t add processing difficulty because in order to make sense of a universal quantifier like each pupil additional contextual restriction to a particular set is required anyway. That is, a sentence like (25-b) even without the partitive is normally not understood as quantifying over all the pupils there have ever been.

The number of characters and words was kept constant across conditions. 96 fillers (48 nonsensical) were added and we constructed two lists in a latin square design.

4.1.2 Procedure

The experiment was a self-paced reading study using moving window presentation (see e.g. Haberlandt, 1994). After each trial, participants had to provide a sensicality judgment within a time limit of five seconds. The experiment began with a practice session of ten trials which was followed by four experimental
blocks with individually randomized order of presentation both within and between blocks. An experimental session took no longer than half an hour.

4.1.3 Participants
16 native German speakers (mean age 22.4 y., range 20-26 y., 13 female) from Tübingen University took part for payment of 5€. None of them had participated in the first two experiments.

4.2 Results
We corrected the reading times for outliers by trimming values that were below 100ms or above 2500ms. This affected less than 1% of the data. Figure 3 depicts the reading times region by region.

At the second quantifier region, the non-bound condition was read numerically even slower than the bound pronoun condition (mean RT: 1244 ms vs. 1159 ms). Paired t-tests revealed that this difference was not statistically reliable ($t_1(15) = -1.55, p = .14$; $t_2(31) = -1.97, p = .06$). Similarly, neither of the following two spillover regions yielded a reliable difference (spillover 1: $t_1(15) = -1.99, p = .34$; $t_2(31) = -1.00, p = .32$; spillover 2: $t_1(15) = -1.65, p = .12$; $t_2(31) = -0.97, p = .34$). Again, second quantifiers with a demonstrative pronoun were read numerically slower than quantifiers with a bound pronoun (spillover 1: 587 ms vs. 555 ms; spillover 2: 1355 ms vs. 1272 ms).

Sensicality ratings revealed that both conditions were judged equally plausible. The bound pronoun condition was rated sensible 81.25% of the time and the non-bound condition was rated sensible 80.86% of the time.
4.3 Discussion

This experiment shows that a bound variable pronoun is not necessarily more difficult than a demonstrative. In sentences with two quantifiers, a bound variable in the restrictor of the second quantifier is as easy to interpret as a demonstrative in the same position. This finding is at odds with some of the online studies on bound pronoun interpretation reported in the beginning of this section. But obviously, the results are hard to compare since we used quantified statements in both the bound variable and in the non-bound condition which the other studies did not. Our findings are compatible with those of Koornneef (2008) who also didn’t find enhanced processing load when readers had to compute bound variable interpretations.

The crucial point for our current purposes is, however, that in the constructions to be used in the next three experiments a bound variable doesn’t increase processing difficulty per se. The next experiment shows that changing the order of quantifiers dramatically changes this picture.

5 Experiment 4a: Computing Inverse Scope in Complete Sentences

After having established all our assumptions, we are now in a position to address the ISH. The present self-paced reading experiment was conducted to find out whether inverting the scope of quantifiers is a cognitively demanding operation.

5.1 Methods

5.1.1 Materials, Predictions and Procedure

We used the 32 items from the previous experiments and constructed them in the four conditions (26-a) to (26-d) (repeated from (4)). Vertical lines indicate segmentation. The last two segments served as spillover regions. They always contained a prepositional phrase which was partitioned into the preposition and the rest of it.

(26) a. Jeden seiner Schüler | lobte | genau ein Lehrer | voller |
    Each of his pupils\textsubscript{acc} praised exactly one teacher\textsubscript{nom} full-of |
    Wohlwollen. goodwill.
    Exactly one teacher praised each of his pupils full of goodwill.

b. Jeden dieser Schüler | lobte | genau ein Lehrer |
    Each of these pupils\textsubscript{acc} praised exactly one teacher\textsubscript{nom} |
    voller | Wohlwollen.
    full-of | goodwill.
    Exactly one teacher praised each of these pupils full of goodwill.
We used a 2×2 factorial design with the two within factors type of second DP (two levels: quantifier (QQ) vs. definite description (QDef)) and type of pronoun (two levels: his vs. these). In addition to the bound variable condition disambiguating towards inverse scope (26-a) and the doubly quantified control in (26-b) we included two non-quantificational control conditions (26-c) and (26-d). In (26-c) his receives an intra-sentence coreferential interpretation (along the lines of this teacher's pupil), which should be as easy as (26-d). The subject DP always appeared at the end of a complete minimal sentence, ie. after the main verb.

In the QQ conditions, we expected increased processing difficulty at the second quantifier if the first quantifier contained a bound pronoun compared to the condition without a possessive in the first quantifier. By contrast, in the QDef conditions the second DP dieser Lehrer should be read equally fast since both his and these are interpreted coreferentially.

In general, we expected doubly quantified sentences to be more difficult than sentences with only one quantifier and a definite description. This prediction immediately follows from the truth conditions of the doubly quantified sentences (ignoring the pronouns for a moment): (27-a) are the truth conditions of the ∀∃! reading, (27-b) for the ∃∀! reading and (27-c) for the QDef sentences. While (27-a) and (27-b) require two conditions, (27-c) is simpler and involves only one condition. This exemplifies that there are considerable differences in semantic complexity between the QQ conditions – irrespective of which scope reading is chosen – and the QDef conditions.

(27) a. For each pupil there is a teacher praising him, but for each there is no more than one.
   b. There is a teacher praising all the pupils, but no other teacher praises all of them.
   c. There is a (unique) teacher praising all the pupils.

The truth conditions are natural language paraphrases of their according first order logic representations. Note that equivalent representations in generalized quantifier theory would have been completely different. It is, however, yet unclear what kind of representations comprehenders use when it comes to understanding quantifiers (see eg. the discussions in Geurts and van der Slik (2005), Hackl (2009), Szymanik (2009) Pietroski et al. (2009) and Lidz et al. (ress)). We think that the first order representations in (27) nicely fit the verification strategies that are actually employed when it comes to evaluating sentences with exactly one.
In addition to the experimental items we included the 96 fillers from the preceding Experiment and constructed four lists using a latin square design. The experimental procedure was the same as in the previous experiment.

5.1.2 Participants

32 native German speakers (mean age 22.5 y., range 18-40 y., 22 female) from Tübingen University took part in the study for the payment of 5€. None of them had participated in the previous experiments.

5.2 Results

We corrected the reading times for outliers by trimming values that were below 100ms or above 2500ms. This affected less than 0.5% of the data. Figure 4 depicts the reading times region by region.

At the verb immediately preceding the second quantifier all conditions were read equally fast. ANOVAs analyzing the RTs of the verb region did not reveal any reliable differences between conditions (all $F_{1/2} < 1$).

At the second quantifier region, the QQ-his condition was read more slowly than the others. It had a mean RT of 1362 ms, whereas the second quantifier in the QQ-these condition was read with a mean RT of 1199 ms. In contrast, the his-control condition had a mean RT of 1171 ms and was even read slightly faster than the these-controls which had a mean RT of 1231 ms. In ANOVAs, this pattern of effects yielded a significant interaction of DP type and pronoun ($F_{1}(1,31) = 11.13; p < .01$; $F_{2}(1,31) = 8.83; p < .01$). Paired t-tests revealed that the QQ-these condition was read reliably faster than QQ-his ($t_{1}(31) = 3.09, p < .05$; $t_{2}(31) = 1.97, p = .06$) but the definite controls did not differ from each other significantly ($t_{1/2} < 1.3$). Neither the main effect of DP type
At the following spillover region, the \textit{QQ-his} condition had a mean RT of 612 ms, the \textit{QQ-these} condition had a mean RT of 578 ms and the two controls had mean RTs of 537 ms and 546 ms, respectively. Statistical analyses revealed a significant main effect of \textit{DP type} \((F_1(1, 31) = 9.72; p < .01; F_2(1, 31) = 13.82; p < .01)\), but neither a reliable main effect of \textit{pronoun} \((F_{1,2} < 1)\) nor a significant interaction \((F_1(1, 31) = 3.33; p = .08; F_2(1, 31) = 1.31; p = .26)\). The effect of \textit{DP type} is due to \textit{QQ} sentences being read more slowly than \textit{QDef} sentences.

The same pattern of RTs showed up even more drastically at the last region. The \textit{QQ-his} condition had a mean RT of 1269 ms, the \textit{QQ-these} condition had a mean RT of 1272 ms and the two controls had mean RTs of 1074 ms and 1066 ms, respectively. ANOVAs revealed a reliable main effect of \textit{DP type} \((F_1(1, 31) = 26.17; p < .01; F_2(1, 31) = 21.87; p < .01)\) reflecting that the \textit{QQ} conditions took longer to read than the \textit{QDef} conditions. Both the main effect of \textit{pronoun} and the interaction were far from significant \((all F_{1/2} < 1)\).

Sensicality ratings showed the same pattern as the reading times. The \textit{QQ-his} condition was rated to be sensible 66.27\% of the time, the \textit{QQ-these} condition 77.25\% of the time and the two controls 90.20\% and 88.98\% of the time. The average judgment RTs were 1441 ms in the \textit{QQ-his} condition and 1514 ms in the \textit{QQ-these} condition. The control conditions were judged faster: \textit{his}-controls had a mean judgment RT of 1302 ms and \textit{these}-controls had a judgment RT of 1249 ms. In ANOVAs analyzing judgment RTs only the main effect of \textit{DP-Type} was reliable \((F_1(1, 31) = 16.23; p < .01; F_2(1, 31) = 12.37; p < .01; all other effects: F_{1/2} < 2.2)\).

\subsection*{5.3 Discussion}

The present experiment shows that forcing inverse scope in doubly quantified sentences leads to considerable processing difficulty. In the \textit{QQ} sentences, the second quantifier took approximately 200 ms longer to read if the first quantifier contained an inversely bound pronoun than when the first quantifier did not contain a pronoun that had to be inversely bound. This effect cannot be due to simple lexical difference between \textit{his} and \textit{these} because the controls did not differ in RT. Furthermore, the same pattern of effects showed up in the sensicality judgments providing further support for the predicted differences in how easy it is to compute a sensible interpretation.

Turning to the \textit{QQ} vs. \textit{QDef} effect, the second prediction was also fully supported by our findings. Irrespective of variable binding, \textit{QQ} sentences were hard to comprehend than \textit{QDef} sentences. Interestingly, this effect had a rather different time course than the scope inversion effect. While the latter effect was limited to the second quantifier, the \textit{QQ} vs. \textit{QDef} effect only showed up one region later and was most pronounced at the end of the sentence. An explanation for this contrast could be to link the scope inversion and the \textit{QQ} vs. \textit{QDef} effects to functionally different aspects of semantic interpretation. The
first could reflect the computation a well-formed scope representation in which all the variables are properly bound, some kind of “logical form”, whereas the second could be due to computing a semantic value for it – i.e. a model fulfilling the truth conditions. We will come back to the question whether the two effects appear in a strict serial order in the eyetracking experiment (Exp. 5).

To conclude, the present experiment established the predicted scope inversion effect. While in Experiment 3 the two QQ conditions with the reversed order of quantifiers did not differ, the bound QQ condition lead to substantial integration difficulty of the second quantifier in the present experiment. Although this effect showed up immediately at the second quantifier region, the present experiment cannot tell us whether scope computation starts exclusively on the basis of the quantificational information, since at this point the processor was already dealing with a complete minimal sentence. This brings us back to the incrementality of semantic interpretation: does the scope inversion effect also show up at the second quantifier when the main verb is still missing? This question was addressed in the next experiment.

6 Experiment 4b: Computing Inverse Scope in Yet Incomplete Sentences

In the present experiment we presented the second quantifier before the verbal predicate. Whereas the ISH predicts a scope inversion effect regardless of the verb position, the Global Interpretation Hypothesis (GIH) predicts no immediate disruption at the second quantifier but only a delayed effect at the point when the sentence is complete.

6.1 Methods

6.1.1 Materials, Predictions and Procedure

We used the 32 items from the previous experiments and constructed them in the four conditions (28-a)–(28-d) (repeated from (5)). All four conditions were identical to those from the last experiment except for the position of the participle gelobt. The second quantifier was separated from the verb by two segments to provide enough space for spillover.

(28) a. Jeden seiner Schüler | hat | genau ein Lehrer | voller |
    Each of-his pupils(acc) | has | exactly one teacher(nom) | full-of |
    Wohlwollen | gelobt. 
    goodwill | praised.

Exactly one teacher praised each of his pupils full of goodwill.

b. Jeden dieser Schüler | hat | genau ein Lehrer | voller |
    Each of-these pupils(acc) | has | exactly one teacher(nom) | full-of |
    Wohlwollen | gelobt. 
    goodwill | praising.
Exactly one teacher praised each of these pupils full of goodwill.

- Jeden seiner Schüler | hat | der Lehrer | voller |
  Wohlwollen | gelobt.
  The teacher praised each of his pupils full of goodwill.

The teacher praised each of his pupils full of goodwill.

We made a small change to the control conditions (28-c) and (28-d) and substituted this N by the N. This was because we had the impression that in the control condition (28-d) a simple definite description makes the sentences more natural.

The findings of the last experiment led us to expect that the QQ conditions should be harder to interpret than the QDef conditions. The resulting difference in RT should show up only after the scope inversion effect. With respect to the scope inversion effect, the ISH and the GIH make fundamentally different predictions. The ISH predicts an immediate inversion effect at the second quantifier just as in the previous experiment whereas the GIH predicts that the processor will have to wait until it has processed the complete sentence.

Again, we included the 96 fillers from the previous experiments and constructed four lists using a latin square design. The procedure was the same as in the previous two experiments.

6.1.2 Participants

32 native German speakers (mean age 23.4 y., range 19-30 y., 25 female) from Tübingen University took part in the study for the payment of 5€. None of them had participated in any of the previous experiments.

6.2 Results

To correct for outliers reading times below 200ms or above 2500ms were trimmed. This correction affected less than 0.3% of the data. Figure 5 depicts the mean reading times region by region.

At the auxiliary directly preceding the second DP, reading times did not differ between conditions (all $F_{1/2} < 1$). At the critical second DP region, the QQ-his condition had a mean RT of 1235 ms which was slightly higher than the QQ-these condition which had a mean RT of 1187 ms. The his- and the these-controls had mean RTs of 818 ms and 864 ms, respectively. ANOVAs revealed a significant main effect of DP type ($F_1(1, 31) = 109.46; p < .01; F_1(1, 31) = 61.25; p < .01$) which was probably due to the lexical differences between exactly one N and the N. Crucially, neither the main effect of pronoun type ($F_{1/2} < 1$)
nor the interaction \( (F_1(1,31) = 3.66; p = .07; F_1(1,31) = .40; p = .53) \) were significant. A paired t-test comparing the two QQ conditions revealed that the numerical difference was far from reliable \( (t_1(31) = .97; p = .33; t_2(31) = .13; p = .90) \). There was no indication of a scope inversion effect at the second quantifier region.

At the following two spillover regions there was also no indication of a scope inversion effect. At the first spillover region the DP type had an influence on the reading times with the QQ conditions being read more slowly than the control conditions. ANOVAs revealed a significant main effect of DP type \( (F_1(1,31) = 109.46; p < .01; F_1(1,31) = 61.25; p < .01) \), but again neither a reliable main effect of pronoun \( (F_{1/2} < 1) \) or a significant interaction \( (F_{1/2} < 1) \). At the second spillover region RTs didn’t differ from each other at all. ANOVAs analyzing the RTs in this region did not reveal any reliable differences \( (all F_{1/2} < 2.3; interaction: F_{1/2} < 1) \). Taken together, up to this point there was no scope inversion effect.

At the verbal predicate, the QQ conditions (mean RT in QQ-his condition 1492 ms; QQ-these condition 1513 ms) were read on average 286 ms slower than the QDef control conditions (his-control 1124 ms; these-control 1218 ms). In ANOVAs, the main effect of DP type was significant \( (F_1(1,31) = 46.56; p < .01; F_1(1,31) = 23.35; p < .01) \), but neither the main effect of pronoun \( (F_{1/2} < 1) \) nor the interaction \( (F_{1/2} < 1) \) were significant. Like in the last experiment, doubly quantified sentences lead to substantially longer reading times at the end of the sentences than sentences with only one quantifier. However, there was no indication of a scope inversion effect on any sentence region.

The sensicality judgments were similar to those in the previous experiment. The QQ-his condition was rated to be sensible 71.09%, the QQ-these condition was rated sensible 76.17% and the two controls 86.33% and 87.11%, respectively. We also analyzed the judgment times to find out whether the scope inversion
effect would show up there. The QQ conditions (mean judgment times: \textit{QQ-his} 1251 ms vs. \textit{QQ-these} 1241 ms) took on average 144 ms longer to judge than the controls (1124 ms and 1082 ms respectively), but there was no difference between the two pronoun types. ANOVAs analyzing the judgment RTs only revealed a significant main effect of \textit{DP type} ($F_1(1,31) = 9.56; p < .01; F_1(1,31) = 10.57; p < .01$; other effects: $F_{1/2} < 1$).

### 6.3 Discussion

In the present experiment we tested whether inverse scope can be computed in yet incomplete sentences before the verbal predicate has been encountered. The ISH predicted enhanced processing difficulty in the \textit{QQ-his} condition immediately at the second quantifier region. Our findings do not support this prediction. The \textit{QQ-his} condition was not harder to comprehend than the \textit{QQ-these} condition neither at the second quantifier nor at the following two spillover regions or in judgment times. In fact, in the present experiment we didn’t find a scope inversion effect at all.

At first sight, our findings are more compatible with the GIH. It predicts that the computation of scope can only start when the sentence is complete and therefore no scope inversion effect is expected to occur at the second quantifier. Instead the scope inversion effect should be delayed until the verbal predicate has been encountered. To our surprise, however, we didn’t find a delayed effect either. How can this total lack of effect be explained? A possible explanation is that participants computed the required scope readings when they encountered the verb but the scope inversion effect might have been covered by the co-occurring stronger QQ vs. QDef effect. This must remain pure speculation, however, because it is impossible to conclude anything from a null effect.

We conducted a final experiment to check whether this line of reasoning is on the right track. We measured eye movements during reading to disentangle early from later effects of scope interpretation. Eyetracking provides us with a more sensitive measure which provides us with a number of measures tapping into different processing stages. These should allow us to gain access to the relative time courses of the scope inversion and the QQ vs. QDef effects. Furthermore, we included both the early verb and the late verb conditions into the same experiment to directly compare the processing of scope in the two word orders.

### 7 Experiment 5: Eyetracking during Reading

Monitoring eye-movements during reading provides a more fine-grained measure of processing difficulty than self-paced reading. With eyetracking it is possible to distinguish early effects reflecting the computation of the initial interpretation from later effects having to do with reanalysis and revisions of the initial interpretation\(^6\). We can, thus, speculate about how a scope inversion effect

\(^6\)We would like to emphasize that this is not meant as taking any particular stand with respect to whether the processor employs only one representation at a time or can activate
could look like. There are a number of theoretical alternatives. All of them are consistent with the reading time data of the just reported self-paced reading experiments. Let’s first consider what to expect in the verb-second word order.

The most “charitable” or “prudent” option is a kind of processor that takes into account the presence of the variable right from the start. Encountering the second quantifier it automatically receives a wide scope interpretation to achieve binding. This can be expected to be more costly than processing the second quantifier in a linear scope configuration because it might require exceptional long movement or type shifting of the second quantifier or whatever the underlying operation may be. In this scenario, inverse scope is the initial interpretation of doubly quantified sentences with a bound variable in the first quantifier; once computed, no scope revision is required. In eyetracking, we would thus expect to find a scope inversion effect at the second quantifier, but the processing difficulty should show up during the initial interpretation, that is it should affect early eyetracking measures and less so regressive eye-movements during re-reading the sentence.

A second alternative is that variable binding does not guide the initial stage of semantic interpretation, but is checked immediately after quantifier scope has been computed. In our configuration of quantifiers, this means that initially the linear scope reading will be computed, but since the variable in the restriction of the first quantifier needs to be bound, scope has to be revised and reanalyzed towards an inverse interpretation. If this explanation is correct, we expect a slightly different pattern of effects. Although a scope inversion effect should also occur at the second quantifier, it should mainly affect regression related eyetracking measures as opposed to measures related to the initial interpretation.

Finally, readers could automatically assign linear scope and notice a binding problem, but delay repair until some later decision point. A natural decision point could, for instance, be the end of the sentence during normal reading. However, the decision point might be shifted to earlier parts of the sentence if the task requires voluntary decisions to move on, as in stops-makes-sense judgments or self-paced reading. It is, thus, well possible that we will find later scope inversion effects in the eyetracking experiment than in Experiment 4. If this theoretical alternative is correct, we would expect to observe scope revision only at the end of the sentence. In addition, it is possible that noticing a binding problem at the second quantifier will slowdown integration of the second quantifier. However, since according to this theoretical alternative scope isn’t recomputed at the second quantifier we would not expect to observe an increase in regressions to the first quantifier phrase.

Turning to the verb-final constructions, it is open what to expect. If self-paced reading in Exp. 4b was too coarse a measure to be able to detect a scope inversion effect at the second quantifier region, we may expect to find the same reading time pattern in both word orders in eyetracking. If, on the other hand, several differently ranked scope alternatives in parallel. We take reanalysis as a shorthand for either revising the first interpretation creating a new one or activating a dispreferred alternative above a certain threshold.
a scope inversion-effect occurred at the last word of the sentence but it was covered by the QQ vs. QDef effect, we may expect the reading time patterns to differ in the two word orders. Whereas the verb-second conditions should lead to an immediate scope inversion effect, the verb-final conditions can be expected to reveal a scope inversion effect that is delayed until the main verb, ie. the end of the sentence.

7.1 Methods

7.1.1 Materials

We used the 36 items from Experiment 1 and added four items to the list of experimental sentences. Each item was constructed in eight conditions according to a 2 (verb position: v-2nd vs. v-final) x 2 (DP type: Q-Def vs. Q-Q) x 2 (pronoun: his vs. these) within design. (29) is a sample item in v-2nd word order and (30) in v-final word order. We adjusted the number of words of the definite descriptions in the QDef controls to the second quantifier phrase in the QQ conditions by adding adjectives to them. Overall, the definite descriptions in the QDef controls were slightly longer than the second quantifier in the QQ conditions (mean length of the second DP: QQ = 19.0 char.; QDef = 20.9 char.; \( t(78) = 2.12; p < .05 \)).

   | | of-his pupils_{acc} | praised | exactly one teacher_{nom} | full-of | goodwill.
   | | | | | | detail.
   Each of his pupils praised exactly one teacher full of goodwill.

   | | of-these pupils_{acc} | praised | exactly one teacher_{nom} | full-of | goodwill.
   | | | | | | detail.
   Each of these pupils praised exactly one teacher full of goodwill.

c. Jeden seiner Schüler | lobte | der nette Lehrer | voller | Wohlwollen.
   | | of-his pupils_{acc} | praised | the nice teacher_{nom} | full-of | goodwill.
   | | | | | | detail.
   The nice teacher praised each of his pupils full of goodwill.

d. Jeden dieser Schüler | lobte | der nette Lehrer | voller | Wohlwollen.
   | | of-these pupils_{acc} | praised | the nice teacher_{nom} | full-of | goodwill.
   | | | | | | detail.
   The nice teacher praised each of these pupils full of goodwill.

   | | of-his pupils_{acc} | has | exactly one teacher_{nom} | full-of | goodwill | praised.
   | | | | | | | detail.
   Each of his pupils has exactly one teacher full of goodwill.
Exactly one teacher praised each of his pupils full of goodwill.

b. Jeden dieser Schüler hat genau ein Lehrer voller Wohlwollen gelobt.
Each of these pupils has exactly one teacher praised.

Exactly one teacher praised each of these pupils full of goodwill.

c. Jeden seiner Schüler hat der nette Lehrer voller Wohlwollen gelobt.
The nice teacher praised each of his pupils full of goodwill.

d. Jeden dieser Schüler hat der nette Lehrer voller Wohlwollen gelobt.
The nice teacher praised each of these pupils full of goodwill.

Regions of interest (ROIs) were defined as in Experiments 4 and 5. Each sentence was paired with the question “was this a sensible sentence?”. The items and 118 fillers (52 fillers were nonsensical, 56 were from an unrelated experiment) were distributed over eight lists in a latin square design.

7.1.2 Participants

48 students from Tübingen University took part in the experiment (mean age 25.3 years, range from 19–35 years; 36 female). They received a payment of 8€. None of them had participated in any of the previous experiments. Six participants were randomly assigned to each list. Six additional participants had to be excluded from the analysis due to calibration problems (N = 4) or error rates above 40% in the practice (N = 2).

7.1.3 Procedure

A desktop-mounted Eyelink 1000 eyetracker monitored the gaze location of participants dominant eye. The eyetracker has a spatial resolution of 0.01 degrees of visual angle and samples gaze location every millisecond. Participants viewed the stimuli binocularly on a 19 inch monitor 70 cm from their eyes. A head rest minimized head movements. The experiment was implemented using the SR Research Experiment Builder software and eyetracking data were exported with the SR Research Data Viewer.

Subjects were tested individually. The tracker was calibrated using a 3x3 grid and we made sure that all fixations were less than 0.5 degrees apart from the calibration stimuli. After calibration was completed, participants read the experimental instructions on the screen. This was followed by a practice session of 10 items. In the experiment, each trial started with a calibration check. The tracker was recalibrated as necessary. Eye-movements were recorded during
reading. In addition, the judgments and the judgment RTs were logged.

The trial began with the presentation of a screen which served as calibration check with a little black dot in the position where the center of the first word would appear. If no fixation was registered within five seconds, recalibration was enforced. Otherwise a sentence appeared in the center of a navy blue screen. It was printed in yellow letters with 15 point font size. Three characters corresponded approximately to one degree of visual angle. After reading the sentence participants had to move their eyes to an asterisk at the bottom of the screen. Fixating the asterisk triggered the presentation of the question screen. In the experimental trials, questions always queried whether the sentence was sensible, but in the filler trials participants also received ordinary comprehension questions. Across the experiment, the ratio between sensicality judgments and ordinary comprehension questions was 2:1. There was no time limit for answering the questions.

7.1.4 Analysis

Prior to all analyses we preprocessed the data. Fixations shorter than 80 ms and within one character space of the previous or next fixation were assimilated to this fixation. The remaining fixations shorter than 80 ms and longer than 1200 ms were excluded. This affected 5.6% of the data. The reading time data were analyzed by computing linear mixed effects model analyses with verb position, DP type and pronoun as fixed effects and participant and item as random effects. This type of analysis is particularly robust to missing values (Baayen et al., 2008), i.e., regions that were skipped. P-values were estimated using Monte-Carlo sampling. To analyze the proportions of regressions we computed logit mixed effects model analyses. To analyze the proportions of “yes, sensible” judgments we computed a Logit Mixed Effects Model and included verb position, DP type and pronoun plus their interactions as fixed effects and the random intercepts of participants and items.

We analyzed fixation times with respect to five eyetracking measures. First-pass time is the total time spent in an interest area before the reader moves on or looks back in the text. If a region of interest was skipped during first-pass, it was counted as a missing value. Go-past times are the sum of fixation durations from the time the reader enters a region, to the time when the reader enters the following region, that is it includes first-pass time plus the time spent on regressions. Go-past times are also often referred to as regression path durations. Again, skipped regions were counted as missing values. Finally, second pass time is the sum of all (re-)fixations on a particular region during regressive eye-movements. If a region received no regressive fixations, its second pass time was set to a value of zero. We also measured two types of proportions of regressions: first-pass regression ratios, i.e., the proportions of how often readers launched a regression from a region during first pass (forward) reading. The proportion of regressions in a region is a measure of how often it was entered from the right.

7In addition to the measures reported here we analyzed first fixation durations. Since there were no differences we refrain from reporting these.
Table 3: Mean proportions of “yes, sensible” judgments (+ mean judgment RTs) in Exp. 6.

<table>
<thead>
<tr>
<th>DP Type</th>
<th>Pronoun</th>
<th>Verb Position</th>
<th>Verb Second</th>
<th>Verb Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>QQ</td>
<td>His</td>
<td>82.5% (1712ms)</td>
<td>81.7% (1690ms)</td>
<td></td>
</tr>
<tr>
<td>QQ</td>
<td>These</td>
<td>84.6% (1732ms)</td>
<td>87.1% (1746ms)</td>
<td></td>
</tr>
<tr>
<td>QDef</td>
<td>His</td>
<td>84.2% (1713ms)</td>
<td>81.7% (1856ms)</td>
<td></td>
</tr>
<tr>
<td>QDef</td>
<td>These</td>
<td>86.7% (1667ms)</td>
<td>82.5% (1677ms)</td>
<td></td>
</tr>
</tbody>
</table>

While first-pass times and first-pass regression ratios reflect early processing stages, second pass times and the proportions of regressions in a region can inform us about revisions of the first interpretation. Go-past times contain both the first-pass times but also time spent during regressive eye-movements. With regards to the targeted processing stage, they are, therefore, somewhere in between early and late measures.

8 Results and Discussion

8.1 Sensicality judgments

Mean judgments and judgment RTs are presented in Table 3. All conditions were overwhelmingly judged to be sensible with a mean of 83.9% “yes, sensible”. This analysis revealed no reliable differences between conditions (all $p > .04$). Judgment RTs were also rather similar across conditions with a mean judgment time of 1724ms. The linear mixed effects model analysis revealed that judgment RTs didn’t differ significantly from each other (all $t < 1.5$).

The lack of differences indicates that participants were able to process the scope inverted QQ-*his* conditions to the same degree as they could process the linear scope QQ-*these* conditions and the (even easier) definite controls. Furthermore, the lack of differences suggests that participants already had finished computing a scope representation when they proceeded to the question screen.

8.2 Eye-Movements during Reading

In the following, we will first report the early eyetracking measures including first-pass times, first-pass regression ratios and go-past times. Then we will look at the proportions of regressions in and second pass times to determine where readers regressed to when they reread earlier parts of the sentences.

8.2.1 Early Eyetracking Measures

Tables 4–6 report the mean first-pass times, first-pass regression ratios and go-past times as well as summaries of the associated statistics. Starting with the pre-critical verb ROI, conditions started to differ. Lexical verbs took longer
to read than auxiliary verbs and verbs preceding an *exactly one* quantifier had slower first-pass times than verbs preceding a definite article. The latter effect of *DP type* is expected if we take into account preview from the verb to the following determiner: the definite article is much more frequent than *exactly* and therefore was presumably accessed faster than the latter. This effect was mediated by the kind of verb. It was stronger for lexical verbs than for auxiliary verbs. This is plausible, since the rather short first-pass times of the auxiliaries, if fixated at all, indicated that the auxiliaries were merely used as landing sites from which readers immediately programmed the next saccade before they had
Table 5: Mean go-past times in ms and linear mixed effects model results for Exp. 6 (*p < .05; **p < .01; ***p < .001).

At the critical second DP first-passtimes were slower in the QQ-his condition than in the QQ-these condition. The interaction of DP type and pronoun looks like the expected scope inversion effect especially since the control conditions showed exactly the opposite pattern. The binding effect was more pronounced in the verb-second conditions where QQ-his was on average about 50ms slower than QQ-these than in the verb-final conditions where we observed only a 30ms difference. To analyze whether both differences were reliable we computed mixed effects models for each word order separately and analyzed the interactions.
of DP type and pronoun, respectively. The interaction was significant in the verb-second conditions (estimate = −71.1; t = 2.01; p < .05), but not in the verb-final conditions (estimate = −52.1; t = 1.47; p = .14). Even though the interaction was not reliable in the verb-final constructions, trends were the same in both word orders. Does the slow-down imply that readers inverted quantifier scope immediately at the second quantifier? The first-pass regression ratios and the go-past times suggest that they did not. Although a yet unbound variable perturbed the reading process it didn’t lead to regressions from the second quantifier or the following region. This was indicated by the analysis of first-pass regression ratios which revealed the critical DP type by pronoun interaction only at the final region of the sentences. Further evidence comes from go-past times which didn’t differ reliably from each other mid-sentence but revealed a strong DP type by pronoun interaction only at the last region.

Interestingly, the eye-movements were rather similar in the verb-second and the verb-final conditions. None of the three eyetracking measures revealed a significant three-way interaction. The only difference consisted in a somewhat weaker DP type by pronoun interaction in the verb-final word order. This is, however, not very surprising given that at the second quantifier readers had seen less of the sentences in the verb-final than in the verb-second conditions.

Here is a summary of the effects discussed so far. We found early effects at Q2 not carrying over to the following region. At first sight this looks like immediate scope inversion, that is, as if readers computed the inverse interpretation right away without ever considering the linear reading. This is, however, only part of the story. Regressive eye-movements originated mainly from the last region. We will argue that the early effects reflect a failed search for a binder, but computation of inverse scope was delayed until the last region. Only then readers established binding under inverse scope. If this is correct, we should expect to observe a massive scope inversion effect during rereading the sentences. In the next section, we therefore turn to the analyses of regressions in and second pass times showing where regressive eye-movements landed and which regions were affected by scope inversion.

### 8.2.2 Regression Related Eyetracking Measures

Tables 7 and 8 present the mean proportions of regressions in and the mean second pass times as well as summaries of the mixed effects model analyses. The second pass times revealed very clear scope inversion effects, i.e. interactions of DP type and pronoun, across the whole sentence with the only exception of the sentence-final ROI. Irrespective of the verb position, readers spent longer rereading the doubly quantified sentences when the first quantifier contained a bound variable than when it did not. Besides this binding effect, the doubly quantified sentences led to substantially more regressions than the QDef controls. In both, proportions of regressions in and second pass times, the main effect of DP type was highly reliable across all regions of interest. Similar to the findings of the previous experiments, this suggests that doubly quantified sentences are much harder to comprehend than simply quantified sentences with
Table 6: Mean first-pass regression ratios in % and logit mixed effects model results for Exp. 6 (*p < .05; **p < .01; ***p < .001).

<table>
<thead>
<tr>
<th>Region of Interest</th>
<th>Verb/Aux</th>
<th>DP 2</th>
<th>Spillover</th>
<th>Last ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verb Second</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ-his</td>
<td>15.0</td>
<td>15.8</td>
<td>9.0</td>
<td>71.0</td>
</tr>
<tr>
<td>QQ-these</td>
<td>12.9</td>
<td>17.5</td>
<td>11.4</td>
<td>64.1</td>
</tr>
<tr>
<td>QDef-his</td>
<td>12.2</td>
<td>15.0</td>
<td>6.6</td>
<td>53.0</td>
</tr>
<tr>
<td>QDef-these</td>
<td>12.5</td>
<td>17.1</td>
<td>10.4</td>
<td>58.8</td>
</tr>
<tr>
<td><strong>Verb Final</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ-his</td>
<td>8.2</td>
<td>18.3</td>
<td>6.0</td>
<td>76.3</td>
</tr>
<tr>
<td>QQ-these</td>
<td>8.3</td>
<td>19.2</td>
<td>7.6</td>
<td>70.3</td>
</tr>
<tr>
<td>QDef-his</td>
<td>6.1</td>
<td>15.5</td>
<td>7.4</td>
<td>65.8</td>
</tr>
<tr>
<td>QDef-these</td>
<td>9.2</td>
<td>16.3</td>
<td>6.3</td>
<td>65.6</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>-2.16</td>
<td>-1.86</td>
<td>-3.04</td>
<td>0.19</td>
</tr>
<tr>
<td>z-value</td>
<td>-9.22 ***</td>
<td>-9.08 ***</td>
<td>-8.36 ***</td>
<td>0.80</td>
</tr>
<tr>
<td><strong>verb position</strong></td>
<td>-0.83</td>
<td>0.04</td>
<td>0.10</td>
<td>0.71</td>
</tr>
<tr>
<td>z-value</td>
<td>-1.921</td>
<td>0.16</td>
<td>0.23</td>
<td>3.31 ***</td>
</tr>
<tr>
<td><strong>DP type</strong></td>
<td>0.26</td>
<td>0.07</td>
<td>0.42</td>
<td>0.99</td>
</tr>
<tr>
<td>z-value</td>
<td>0.917</td>
<td>0.27</td>
<td>0.99</td>
<td>4.55 ***</td>
</tr>
<tr>
<td><strong>pronoun</strong></td>
<td>0.04</td>
<td>0.16</td>
<td>0.58</td>
<td>0.31</td>
</tr>
<tr>
<td>z-value</td>
<td>0.13</td>
<td>0.64</td>
<td>1.35</td>
<td>1.48</td>
</tr>
<tr>
<td><strong>verb position*DP type</strong></td>
<td>0.10</td>
<td>0.14</td>
<td>-0.63</td>
<td>-0.35</td>
</tr>
<tr>
<td>z-value</td>
<td>0.19</td>
<td>0.40</td>
<td>-1.01</td>
<td>-1.10</td>
</tr>
<tr>
<td><strong>verb position*pronoun</strong></td>
<td>0.46</td>
<td>-0.10</td>
<td>-0.74</td>
<td>-0.33</td>
</tr>
<tr>
<td>z-value</td>
<td>0.82</td>
<td>-0.27</td>
<td>-1.20</td>
<td>-1.08</td>
</tr>
<tr>
<td><strong>DP type*pronoun</strong></td>
<td>-0.24</td>
<td>-0.03</td>
<td>-0.29</td>
<td>-0.72</td>
</tr>
<tr>
<td>z-value</td>
<td>-0.59</td>
<td>-0.09</td>
<td>-0.51</td>
<td>-2.36 *</td>
</tr>
<tr>
<td><strong>verb position<em>DP type</em>pronoun</strong></td>
<td>-0.27</td>
<td>0.05</td>
<td>0.70</td>
<td>0.39</td>
</tr>
<tr>
<td>z-value</td>
<td>-0.36</td>
<td>0.07</td>
<td>0.82</td>
<td>0.88</td>
</tr>
</tbody>
</table>

a definite description instead of the second quantifier. Interestingly, these differences only occur during rereading the sentences. This is suggested by the fact that the early measures didn’t yield a significant effect of DP type for any of the mid-sentence regions.

What does this mean for the incrementality of scope assignment? As it looks, the scope-inversion and the QQ vs. QDef effects occur rather late. The most striking finding in this respect is that the verb position didn’t have an influence on the time course of scope resolution. Even though readers had processed a complete minimal sentence when they encountered the second quantifier in the
verb-second constructions, they waited until the end of the sentence until they seemed to compute relative scope. This finding is clearly incompatible with the Incremental Scope Assignment Hypothesis but fits the Global Interpretation Hypothesis according to which quantifier scope is only determined at the end of the sentence.

If inverse scope is computed only at the end of the sentence how should we interpret the DP type by pronoun interaction in the first-pass times, then? In principle, there are two scenarios which would let us expect this kind of effect.

Scope could be underspecified until the sentence boundary along the lines

<table>
<thead>
<tr>
<th>Region of Interest</th>
<th>QP1</th>
<th>Verb/Aux</th>
<th>DP 2</th>
<th>Spillover</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verb Second</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ-his</td>
<td>78.8</td>
<td>34.2</td>
<td>28.8</td>
<td>32.0</td>
</tr>
<tr>
<td>QQ-these</td>
<td>75.8</td>
<td>29.2</td>
<td>24.2</td>
<td>27.8</td>
</tr>
<tr>
<td>QDef-his</td>
<td>68.3</td>
<td>25.7</td>
<td>20.0</td>
<td>22.3</td>
</tr>
<tr>
<td>QDef-these</td>
<td>73.8</td>
<td>24.6</td>
<td>22.1</td>
<td>27.4</td>
</tr>
<tr>
<td><strong>Verb Final</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQ-his</td>
<td>84.2</td>
<td>35.8</td>
<td>33.8</td>
<td>35.2</td>
</tr>
<tr>
<td>QQ-these</td>
<td>79.6</td>
<td>29.5</td>
<td>33.5</td>
<td>21.5</td>
</tr>
<tr>
<td>QDef-his</td>
<td>74.2</td>
<td>26.7</td>
<td>23.8</td>
<td>20.6</td>
</tr>
<tr>
<td>QDef-these</td>
<td>75.0</td>
<td>23.9</td>
<td>23.8</td>
<td>24.1</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>1.03</td>
<td>-1.17</td>
<td>-1.60</td>
<td>-1.40</td>
</tr>
<tr>
<td>z-value</td>
<td>4.28***</td>
<td>-6.28***</td>
<td>-7.48***</td>
<td>-5.83***</td>
</tr>
<tr>
<td><strong>verb position</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>0.36</td>
<td>0.04</td>
<td>0.25</td>
<td>-0.06</td>
</tr>
<tr>
<td>z-value</td>
<td>1.60</td>
<td>0.17</td>
<td>1.09</td>
<td>-0.21</td>
</tr>
<tr>
<td><strong>DP type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>0.68</td>
<td>0.43</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td>z-value</td>
<td>2.92**</td>
<td>2.05*</td>
<td>2.39*</td>
<td>2.20*</td>
</tr>
<tr>
<td><strong>pronoun</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>0.34</td>
<td>-0.07</td>
<td>0.14</td>
<td>0.35</td>
</tr>
<tr>
<td>z-value</td>
<td>1.51</td>
<td>-0.30</td>
<td>0.59</td>
<td>1.28</td>
</tr>
<tr>
<td><strong>verb position*DP type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>0.07</td>
<td>0.04</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>z-value</td>
<td>0.21</td>
<td>0.11</td>
<td>0.05</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>verb position*pronoun</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>-0.28</td>
<td>-0.079</td>
<td>-0.15</td>
<td>-0.25</td>
</tr>
<tr>
<td>z-value</td>
<td>-0.88</td>
<td>-0.22</td>
<td>-0.45</td>
<td>-0.66</td>
</tr>
<tr>
<td><strong>DP type*pronoun</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>-0.54</td>
<td>-0.17</td>
<td>-0.41</td>
<td>-0.62</td>
</tr>
<tr>
<td>z-value</td>
<td>-1.65</td>
<td>-0.58</td>
<td>-1.27</td>
<td>-1.67</td>
</tr>
<tr>
<td><strong>verb position<em>DP type</em>pronoun</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>estimate</td>
<td>0.12</td>
<td>-0.01</td>
<td>0.40</td>
<td>-0.24</td>
</tr>
<tr>
<td>z-value</td>
<td>0.24</td>
<td>-0.02</td>
<td>0.90</td>
<td>-0.45</td>
</tr>
</tbody>
</table>

Table 7: Mean proportions of regressions in % and logit mixed effects model results for Exp. 6 (*p < .05; **p < .01; ***p < .001).
Table 8: Mean second pass times in ms and linear mixed effects model results for Exp. 6 (*p < .05; **p < .01; ***p < .001).

of semantic underspecification accounts (cf. Lashawi and Crouch (1992) and Reyle (1993), for a recent review see Egg (ress); see also Sanford and Sturt (2002) for a psycholinguistic perspective). The prolonged reading times in the QQ-his conditions could then be due to adding a binding constraint that the first quantifier has to be interpreted in the scope of the second one. The addition of the binding constraint can be expected to consume time and this may explain the early effect. Crucially, we have to assume that at this stage the representation of scope is underspecified and will only be fully determined at the end of the sentence, even though there is only one possible specification. It is exactly the
specification stage that is targeted by the late effects and we only have to assume that specifying an inverse interpretation is more difficult than specifying surface scope.

In the second kind of scenario we have to differentiate between three steps in computing inverse scope. The first consists in the computation of the preferred logical form of the sentence, that is the surface scope representation. This happens immediately when the second quantifier is encountered resulting in the computation of a surface scope representation with an unbound variable in the first quantifier. A binding check reveals that the computed logical form is not well-formed. This failed check is reflected in the first-pass effect. Scope inversion takes place only at the end of the sentence, at the second stage which consists in revising LF towards inverse scope. A binding check then reveals that the revised representation is well formed. The last stage consists in computing a mental model that makes the sentence true. Like the revision of logical form the last step is also carried out only at the processing domain boundary, that is the end of the sentence. This explains why we see a general increase in difficulty of doubly quantified sentences as compared to the QDef controls at the end of the sentence. It is important to stress that this increase was observable even in the surface scope QQ-these conditions. It indicates that even in these conditions some aspect of scope computation were delayed until the clause boundary. The additional increase in difficulty in the QQ-his conditions is also accounted for. In the latter conditions scope is revised during the second stage, whereas the QQ-these conditions do not require recomputation of scope.

Obviously, our data are consistent with both of these two theoretical alternatives. However, no matter which of the two alternatives is correct none involves a fully specified representation of relative scope until the end of the sentence. So, even if they look rather different at first sight they share a lot once we take a closer look.

9 General Discussion

A short summary of the findings may be in order before we synthesize the results of the six experiments and put them in a larger perspective. The present study investigated whether quantifier scope, a classic source of semantic ambiguity, is computed immediately as soon as a quantifying expression is encountered or is interpreted more globally after a complete sentence. We measured reading times of sentences in which variable binding excludes the preferred scope interpretation and instead disambiguates the sentence towards a dispreferred interpretation, yielding a kind of semantic garden-path.

The first three experiments were conducted to establish three underlying assumptions about the interpretation of these constructions. In the first experiment we gathered scope judgments in two object topicalized constructions with different order of quantifiers. Sentences with the determiner each of these preceding an exactly one quantifier turned out to meet the requirements of the online experiments: the linear interpretation was strongly preferred over the
inverse reading, but the latter was still available. The second experiment investigated how the presence of a possessive pronoun in the first quantifier affected the available readings of this construction. The paraphrase selection task revealed that a possessive pronoun in the first quantifier overwhelmingly led to a bound variable interpretation with inverse scope. We thus observed the expected shift from linear to inverse readings due to the presence of a variable. Furthermore, it showed that comprehenders chose bound variable interpretations even if a discourse referent was introduced in the sentence that could serve as an antecedent for a coreferential interpretation. This finding makes it unlikely that variable binding is a last resort readers will only consider when no other option is available. The third experiment was a self-paced reading study which showed that a bound variable interpretation isn’t difficult per se. If the variable appeared in a quantifier which followed its binder, no difficulty emerged. We can thus be sure that any difficulty that arises in doubly quantified sentences with a variable in the restriction of the first quantifier are due to scope inversion and not to difficulty of binding (or co-indexing) per se.

The next two experiments tested scope inversion effects in sentences in which we manipulated the position of the lexical verb. Experiment 4a tested sentences with the second quantifier following the main verb. So, at the point when readers encountered the second quantifier they already had seen the entire sentence. This was different in Experiment 4b where the lexical verb appeared three segments after the second quantifier in sentence final position. The doubly quantified sentences were compared to QDef controls which had a definite description instead of a second quantifier. In both experiments we observed a general slowdown in reading time of the doubly quantified sentences compared to the QDef controls at the sentence final region regardless of whether the first quantifier contained a variable or not. The expected scope inversion effect due to variable binding, however, could only be detected in Experiment 4a where we found clear differences in the reading times of the second quantifier, whereas Experiment 4b did not reveal a QQ-his vs. QQ-these effect at any sentence region. The final experiment investigated eye-movements during reading of all eight conditions from the previous two self-paced reading experiments. Against the Incremental Scope Assignment Hypothesis we found a scope inversion effect that was consistent across eyetracking measures only at the very end of the sentences and during rereading them. Verb position had no effect suggesting that in both word orders comprehenders waited until the last region of the sentence before they started to compute (inverse) scope. This is remarkable since in the verb second conditions at the second quantifier the processor had already interpreted a complete minimal sentence. At the second quantifier region we observed a slowdown in the bound variable conditions during first-pass reading, but it didn’t lead to regressive eye-movements. We took this early effect as an indication of a failed binding check which, however, didn’t trigger immediate scope inversion. Instead, scope interpretation was delayed in line with the Global Interpretation Hypothesis:

**Global Interpretation (GIH):** When interpreting a quantifier the processor
waits until both the restriction and the nuclear scope have been fully
determined, that is the computation of scope will only start when the
sentence is complete.

On closer inspection there were puzzling discrepancies between the eyetrack-
ing study and the self-paced reading experiments. We were puzzled first that in
self-paced reading in Exp. 4 the scope inversion effect showed up at the second
quantifier but wasn’t detectable afterwards, whereas in eyetracking it appeared
only at the end of the sentence. The reading time data can be reconciled, though,
if we take into account that in the self-paced reading experiments readers may
have adopted a certain strategy. They may not have moved on to the next
segment before they fully and successfully had interpreted the input at hand. If
this were the case, effects at earlier parts of the sentences may be expected in
self-paced reading than in eyetracking. This is a likely strategy because once a
reader has pressed the button to see the next segment there is no chance of going
back to earlier parts of the sentence. Strategic reading is particularly likely to
occur in sentences that are difficult to process which arguably was the case in
the doubly quantified constructions. Related to this issue, we observed much
higher sensicality ratings for the same doubly quantified sentences in the eye-
tracking study than in the self-paced reading experiments. This indicates that
without having the chance to reread doubly quantified sentences comprehenders
often fail to arrive at a successful interpretation.

Does this mean that scope ambiguity is a phenomenon that is too complex
to still be considered within the realm of language processing at all? Wouldn’t
it be more appropriate to treat scope computation as a kind of problem solving
which is only carried out post-interpretively when an explicit decision about the
reading and/or sensicality is required? Such an account would fit the results
of the ERP study of Dvivedi et al. (2010) who observed a slow negative wave
at a disambiguation following scope ambiguous sentences (as compared to un-
ambiguous controls) irrespective of the reading they were disambiguated to (cf.
Sanford and Sturt (2002) for a similar proposal). It thus looks like if readers
begin with the computation of scope only when they are forced to and they
do so by attentionally reflecting upon the meaning. Under “normal” circum-
stances, however, they do not compute relative scope. Further evidence for the
involvement of decision processes comes from a recent fMRI study by McMillan
et al. (2011) who report recruitment of brain networks associated with strate-
gic decision making in suppressing implausible scope interpretations. Although
an underspecification account seems plausible on the basis of the reading time
data of the present study it doesn’t fit our findings in their entirety.\textsuperscript{8} The first
thing to note is that the scope judgments of Experiment 1 revealed clear differ-
ences in preference and that the two constellations of quantifiers, although
syntactically identical, clearly differed in their scopal preferences. According

\textsuperscript{8}Here, we only consider underspecification accounts of the type discussed in the litera-
ture. As outlined in the last section, we think that an underspecification theory that takes
into consideration the processing domain of quantifier scope and assumes specification at the
domain-boundary would be well suited to model the findings of the present study.
to underspecification accounts it would be expected that in each construction comprehenders can equally easily specify both scope readings, a prediction that is clearly not borne out. Similarly, taking underspecification for granted, it remains completely unclear why we should see a scope inversion effect at all. A particularly strong case against underspecification comes from the paraphrase selection task (Exp. 2) where we tested the doubly quantified sentences with the possessive pronoun in the first quantifier in embedding sentences which explicitly provided a referent for a coreferential interpretation. If comprehenders really avoid scope computation – and hence binding – we would expect to see no bound interpretations in these examples, but, clearly this isn’t what we found. Instead, in a third of all trials and consistently across participants and items, participants computed inverse scope interpretations in choosing a bound variable interpretations. Why should they do so, if computing (inverse) scope was really that difficult and there was an alternative, easier interpretation available to them?

We therefore think of a different explanation. The key concept is the notion of processing domain. It defines the decision point at which a particular interpretation is chosen. Crossing this point will automatically lead to a specified interpretation, but before the processing domain is complete the processor doesn’t decide on a particular interpretation. In the case of relative scope the processing domain is rather large, i.e. the complete sentence. The scope decision is influenced by a number of factors like the lexical semantic properties of the quantifiers (e.g. whether it is distributive or not), the order of quantifiers and others. As a consequence when reaching this decision point the processor may at first arrive at an interpretation that is inappropriate due to binding problems or implausibility. We take this to be the reason for the scope inversion effects starting from the last region of the sentence. Furthermore, the assumption of a sentence-size processing domain for quantifier interaction provides a cue what kind of processes contribute to the well-known (but poorly understood) phenomenon of sentence wrap-up commonly observed in reading time studies (see e.g. Haberlandt (1994)). In our study it didn’t matter what information was specifically introduced by the last region of the sentence (lexical verb or adverbial), the only thing that seemed to matter was that it concluded the sentence.

How do the findings of this paper relate to processing studies of semantic phenomena beyond quantifier scope ambiguities? There is broad consensus that the semantic representation is constructed in close temporal contiguity to the input signal. The most rigorous formulation has perhaps been provided in Hagoort (2003, S22)’s *Immediacy Assumption*: “different information types (lexical, syntactic, phonological, pragmatic) are processed in parallel and influence the interpretation process incrementally, that is, as soon as the relevant pieces of information are available”. The findings of the present study are at odds with this assumption. We believe that instead of addressing the issue of incremental semantic interpretation in general it is important to consider it phenomenon by phenomenon. Besides scope, we have tested how much information is needed before the processor can detect aspectual mismatch and carry out aspectual coercion (Bott, pear). The interpretation of lexical aspect is another example
where semanticists have argued that it is a property of complete predications and not of lexical verbs. The reading time data revealed mismatch and coercion effects only after a complete verb argument structure, so this may be another phenomenon where we have to assume a domain size bigger than a single word. To be clear, this by no means implicates that all semantic phenomena depend on complete sentences. Impressive evidence to the contrary comes from a huge number of online studies.

**References**


