

The manner and time course of updating quantifier scope representations in discourse

Jakub Dotlačil*and Adrian Brasoveanu†

April 21, 2014

Abstract

We present the results of two experiments, an eye-tracking study and a follow-up self-paced reading study, investigating the interpretation of quantifier scope in sentences with three quantifiers: two indefinites in subject and object positions and a universal distributive quantifier in adjunct position. In addition to the fact that such three-way scope interactions have not been experimentally investigated before, they enable us to distinguish between different theories of quantifier scope interpretation in ways that are not possible when only simpler, two-way interactions are considered. The experiments show that contrary to underspecification theories of scope, a totally ordered scope-hierarchy representation is maintained and modified across sentences and this scope representation cannot be reduced to the truth-conditional/mental model representation of sentential meaning. The experiments also show that the processor uses scope-disambiguating information as early as possible to (re)analyze scope representation.

*Center for Language and Cognition, University of Groningen, Groningen, The Netherlands

†Linguistics Department, University of California, Santa Cruz, USA

Keywords: quantifier scope; disambiguation of quantifier scope; reanalysis of interpretation; syntax-semantics interface

Introduction

Sentences with multiple quantifiers (expressions like *every cake*, *each man*, *a boy*) are ambiguous in ways in which sentences with one quantifier are not. Consider, for example, (1). (1a) makes a statement about one individual boy, introduced by the quantifier *a boy*. In contrast to that, (1b) might, but does not have to, make a statement about many boys – one for each girl – because the quantifier *a boy* now interacts with the other quantifier present in the clause, *each girl*.

- (1) a. Mary kissed a boy.
b. Each girl kissed a boy.

More technically, we say that (1b) receives two different interpretations depending on which quantifier *takes scope* over the other. If *each girl* takes scope over *a boy*, the resulting interpretation is: each girl is such that she kissed some boy or other; the boys are possibly different from girl to girl. If *a boy* takes scope over *each girl*, the resulting interpretation is: there is one boy such that each girl kissed him; all the girls kissed the same boy.

Sentences like (1) with multiple quantifiers did not receive a general account in logic until Frege (1879). In psychology, multiple quantifiers are challenging for theories of human reasoning (see, e.g., Johnson-Laird et al. 1989). Finally, psycholinguistic and linguistic research has showed that enumerating the conditions under which quantifiers interact is far from trivial. This research has uncovered, for example, that the order in which quantifiers appear in a sentence matters (Ioup, 1975; Fodor, 1982, a.o.). But order is not the only factor that governs the

interaction of quantifiers. Rich psycholinguistic research on scope (Ioup 1975; VanLehn 1978; Gil 1982; Gillen 1991; Kurtzman and MacDonald 1993; Tunstall 1998; Anderson 2004; Filik et al. 2004; Paterson et al. 2008; Dwivedi et al. 2010; Radó and Bott 2012) has gathered a fairly detailed understanding of what influences interpretation choices that readers/hearers make in interpreting sentences with two quantifiers. These findings have formed the basis of several theories of quantifier scope, and its processing and reanalysis. The main goal of this paper is to argue that this broad range of theories can be narrowed down when we go beyond sentences like (1b) and consider the interaction of three quantifiers.

In the next section, we briefly review the most important semantic notions and findings related to quantifier scope. Afterwards, we show how studying the processing of sentences with three quantifiers can give us an argument for particular theories of scope, its processing and reanalysis. The experimental part discusses two experiments (an eye-tracking and a self-paced reading study) that provide the necessary evidence.

Background: quantifier scope

How can the interpreter identify the scope of a quantifier? Various possibilities were considered in linguistics. One option is that the scope of a quantifier follows from the grammatical-role hierarchy (Ioup, 1975) or from its syntactic hierarchy counterpart (Reinhart, 1983; May, 1985; Heim and Kratzer, 1998): a quantifier taking a particular role scopes over every argument which appears to the right of that role in the hierarchy shown in (2). That is, the quantifier which functions as an Object has an Adjunct quantifier in its scope but not a Subject quantifier, while a Subject quantifier has every argument in its scope.

- (2) Grammatical-Role/Syntactic hierarchy:
Subject \gg Object \gg Adjunct \gg ...

Another alternative is that the scope of a quantifier is decided based on the thematic-role hierarchy (Jackendoff, 1972; Kurtzman and MacDonald, 1993). Yet another alternative is that linear order resolves scope. In that case, the scope of a quantifier is equivalent to the material that appears to the right of the quantifier (Johnson-Laird, 1969; Fodor, 1982; Johnson-Laird et al., 1989).

As has been known at least since Montague (1973), considering only one hierarchy is not sufficient to derive all the available interpretations for sentences with multiple quantifiers. For example, in (3), *every night* is an adjunct following the rest of the clause. Thus, from the perspective of several hierarchies, it should not take scope over the subject or the object. But it is possible, albeit dispreferred, to understand (3) as introducing multiple caregivers into discourse, one per night, as if the indefinite was interpreted in the scope of the distributive quantifier (Ioup 1975; Gil 1982; Tunstall 1998; Anderson 2004; AnderBois et al. 2012, a.o.).

- (3) A caregiver comforted Mary every night.

In order to derive the marked reading of (3), it is often postulated that any hierarchy must have a certain amount of flexibility in it. For instance, much of the linguistic literature on the syntax and semantics of quantifier scope postulates that in a range of cases, Adjuncts can be promoted to the first position in the hierarchy, thereby taking scope over Subjects. This promotion is labeled *inverse scope* because the covert scope representation is the inverse of the overt/‘surface’ one.

- (4) Marked hierarchy for a particular instantiation of inverse scope:
Adjunct \gg Subject \gg Object \gg ...

With this background in mind, we now turn to the discussion of the processing of quantifier scope. Suppose that (3) above is followed by another sentence

that starts either as shown in (5a), or as in (5b). It has been noted repeatedly (Kurtzman and MacDonald, 1993; Tunstall, 1998; Anderson, 2004; Filik et al., 2004; Paterson et al., 2008), that (5a) leads to processing difficulties when compared to (5b) (observed, for example, in increased reading times and regressions in eye-tracking studies).

- (5) a. The caregivers...
- b. The caregiver...

There are currently several theories that can explain this state of affairs. First, the processing difficulties are often (but not always) taken as evidence that the processor postulated a scope representation which is incompatible with the incoming information and this scope representation has to be amended/reanalyzed (Kurtzman and MacDonald, 1993; Tunstall, 1998; Anderson, 2004; Pylkkänen and McElree, 2006).

The reasoning goes as follows. The processor considers the basic hierarchy in (2) by default. However, when the processor encounters (5a), it has to switch from that default scope representation to the marked hierarchy in (4), which allows the subject indefinite to introduce multiple individuals. This reanalysis, as is often the case, incurs processing cost.

There are other possible explanations for the contrast between the processing of (5a) and the processing of (5b). As Filik et al. (2004) note, the processing cost in (5a) might arise simply because readers prefer to match morphological features of the definite to its antecedent noun phrase, which also appears in singular (see also Kemtes and Kemper 1999).

Another possibility is that it is not the reanalysis of scope hierarchy that is costly but scope *specification*, as proponents of underspecification theories of scope would probably argue. In such theories (see Ebert 2005 for a summary and references) one assumes no scopal ordering unless specifically required. If

we abbreviate the underspecified scope as ‘ \approx ’ (which really stands for “ \gg or \ll ”), we can represent the default scope order as shown below:

(6) Subject \approx Object \approx Adjunct

That is, by default, any argument can be in the scope of or take scope over any other argument. This underspecified scope order is compatible with the continuation (5b). In contrast to that, (5a) signals that the Adjunct quantifier must scope over the Subject quantifier, that is, (5a) excludes the option Subject \gg Adjunct because the Subject indefinite could not introduce multiple caregivers in that scope order. Thus, (5a) leads to the following hierarchy:

(7) Adjunct \gg Subject \approx Object

The new hierarchy specifies relations which were originally unspecified. This specification requires extra work on the part of the processor and might be the cause of increased latencies and regressions.

Finally, yet another possibility is that the processor does not incur processing cost because it changes the scope representation of (3), but because it directly modifies the interpretation. For concreteness, let’s assume that the interpretation is a mental model along the lines of [Johnson-Laird \(1969\)](#). [Johnson-Laird et al. \(1989\)](#) propose that distributive quantifiers create a loop, which can be summarized as follows:

- i. introduce one element from the domain of the quantifier
- ii. interpret the scope of the quantifier, introducing a single new element for every indefinite
- iii. go back to step (i.)
- iv. repeat until all elements in the domain of the distributive quantifier are introduced

Assuming that readers start with the default model created by the linear order of quantifiers (Johnson-Laird et al., 1989) and directly modify their mental model only if necessary, they should consider the subject indefinite outside the loop of the distributive quantifier. However, encountering the plural definite *the caregivers* in (5a) should force them to place the indefinite *a caregiver* from the first sentence inside the loop of the distributive quantifier, to satisfy the plurality requirement. This change in their mental model should be reflected in increased processing costs relative to (5b), where no change in the mental model is necessary since (5b) is compatible with the original default interpretation.

In sum, there currently are several distinct theories that can explain both the interpretation possibilities associated with sentences with two quantifiers, and why disambiguating information in the form of a plural definite referring back to an indefinite (as in (5a)) is costly relative to a singular definite (as in (5b)). We believe that it is hard, and maybe even impossible, to decide between these theories when focusing on sentences with two quantifiers. However, the situation changes dramatically once we consider more than two quantifiers.

Current study

As mentioned, scope preferences could be accounted for by, among others, the grammatical-role, thematic-role or linear-order hierarchies. Although these hierarchies differ in details, they have something in common. They are all (strict) total orders, in other words, they satisfy the properties of asymmetry, totality and transitivity. For our purposes, only the last two properties are relevant, so we state them explicitly below:

- (8) a. Totality of scope: For any A, B, either A scopes over B or B scopes over A.

- b. Transitivity of scope: For any A, B and C, if A scopes over B and B scopes over C then A scopes over C as well.

By way of example, consider the hierarchy in (10a) below. This hierarchy is total: for any two elements, it holds that one scopes over the other. Also, it satisfies transitivity. For example, Object has Adjunct in its scope. Subject, in turn, scopes over Object and also *all the other quantifiers* that Object has in its scope.

In light of these observations, we will consider combinations of two indefinites and one universal distributive quantifier, exemplified in (9) below.

- (9) A caregiver comforted a child every night.

The adjunct quantifier *every night* might not take either indefinite in its scope (the default option). Alternatively, it might scope over the object indefinite or the subject indefinite. We can illustrate all three options in terms of the grammatical-role hierarchy: the default surface reading is provided in (10a), the one in which *every night* scopes over the object is provided in (10b), and finally the one in which the universal scopes over the subject is in (10c). Note a fact that will be crucial: when Adjunct scopes over Subject, Object ends up in its scope as well.

- (10) a. Subject \gg Object \gg Adjunct
b. Subject \gg Adjunct \gg Object
c. Adjunct \gg Subject \gg Object

Suppose that as before, (9) above is followed by another sentence that starts either as in (11a) or in (11b). As we noted previously, (11a) should lead to processing difficulties when compared to (11b), which can be explained in various ways, ranging from morphological mismatch to the reanalysis of mental models.

(11) a. The caregivers...

b. The caregiver...

Let's assume for a moment that (11a) is costly because it requires the re-analysis of the scopal representation the processor constructed for sentence (9). In particular, (11a) requires the processor to switch from (10a) to (10c). This in turn makes an interesting prediction: readers should expect the direct object indefinite *a child* to also be interpreted in the scope of *every night*. In fact, this should now be their default interpretation, even though it is *not* the default interpretation of (9).

Concretely, (12a) below should be no harder to process than (12b); in fact (12a) might even be preferred over (12b).

(12) a. ... wanted the children to get some rest.

b. ... wanted the child to get some rest.

On the other hand, if the sentence following (9) begins as in (11b), we expect that (12b) should be the preferred continuation and (12a) should be dispreferred since the default scope representation in (10a) is maintained throughout.

We expect this pattern to emerge only if (11a) – as compared to (11b) – is used to *reanalyze* a total and transitive scopal representation that the processor builds. The other theories of quantifier scope processing discussed in the previous section do not make this prediction.

First, if Filik et al. (2004) were right that (11a) is costly simply because of the morphological mismatch between the plural definite and its antecedent, we would expect (12a) to be always more difficult than (12b), for the same reason. Thus, this theory predicts no interaction between the two continuations in (11) and (12).

No interaction is predicted by underspecification theories of scope either. In their case, (11a) leads to the following specification:

(13) Adjunct \gg Subject \approx Object

Crucially, resolving the scope between Adjunct and Subject, necessary for (11a), leaves the scope interpretation of Object with respect to Adjunct and Subject intact. In such a theory, it is not predicted that the scope resolved by (11a) should affect (12) in any way.

The same holds for the change in interpretation formulated in terms of mental models. This is because the mental models considered by Johnson-Laird et al. (1989) – or the meaning representations usually proposed in formal semantics – do not record any information about the original meaning construction process, they only store the end product. After reading (11a), nothing in the mental model itself forces one to reconsider the scope of the second indefinite. Thus, the plural object in (12a) should cause processing difficulties (when compared to the singular object in (12b)) irrespective of whether the subject definite is singular or plural.

In sum, we see that the totality and transitivity of scope, (8), makes specific predictions about the interaction of three quantifiers, and these predictions are not shared by (i) theories of scope that circumvent totality and transitivity, or (ii) theories that do not assume that processing is sensitive to such reanalysis. The predictions are summarized in Table 1.

[Table 1 about here.]

We now discuss two experiments that enable us to empirically evaluate these predictions.

Experiment 1: An eye-tracking study

Introduction

The experiment investigated the real-time resolution of the scope of indefinites and distributive quantifiers using eye tracking. There were four conditions that disambiguated scope late by manipulating number morphology (singular vs plural) on anaphoric definites as indicated above, and two conditions that disambiguated scope early. We first describe the late-disambiguation conditions.

Each item consisted of two sentences. The first sentence introduced two indefinites, one in the subject position and one in the object position, and a distributive quantifier as an adjoined adverbial modifier following the object. The first sentence was identical across the four conditions. The second sentence, which was different for each condition, provided further information about the scope relations between the distributive quantifier and the indefinites. The four conditions are exemplified in (14) below.

- (14) A caregiver comforted a child every night.
- a. The caregivers[**pl**] wanted the children[**pl**] to get some rest.
(SUBJECT:PL, OBJECT:PL)
 - b. The caregivers[**pl**] wanted the child[**sg**] to get some rest.
(SUBJECT:PL, OBJECT:SG)
 - c. The caregiver[**sg**] wanted the children[**pl**] to get some rest.
(SUBJECT:SG, OBJECT:PL)
 - d. The caregiver[**sg**] wanted the child[**sg**] to get some rest.
(SUBJECT:SG, OBJECT:SG)

Scope relations were disambiguated by means of the number marking on the subject or the object definites in the second (disambiguating) sentence.

One possibility was that both indefinites were assigned narrow scope with respect to the distributive quantifier, which is necessary to license plural morphology on both arguments in the second sentence, (14a). We are going to mark this situation as Subject:Pl, Object:Pl.

The second possibility is shown in (14b) and we will mark it as Subject:Pl, Object:Sg. In this case, the subject had narrow scope with respect to the distributive quantifier and the scope of the object remained ambiguous. We take the singular object to be ambiguous and not necessarily wide scope for a semantic reason. The reason is that a singular entity can still be associated with a narrow scope existential because of so-called ‘accidental co-reference’ situations. For example, when we interpret the indefinite in the sentence *Jane comforted a child every night* as taking narrow scope relative to the universal, the resulting reading of the sentence is that every night, Jane comforted a *possibly different* child. The child can vary from night to night, but this variation is not required. We *allow* the indefinite to co-vary with the universal, but it might accidentally turn out that it takes the same value over and over again – hence the label of ‘accidental co-reference’.

In the third case, Subject:Sg, Object:Pl, exemplified in (14c), the indefinite object had narrow scope with respect to the distributive quantifier, signalled by the plural morphology on the definite object. Finally, Subject:Sg, Object:Sg, (14d), in which both definites appeared in singular, represents a situation in which neither scope was disambiguated.

Besides these four conditions, the experiment included two other conditions in which the first sentence signaled explicitly that one of the two indefinites had narrow scope. The disambiguation was done in two steps. First, the distributive quantifier appeared sentence initially (in topic position), which strongly biased it towards wide scope. Second, we disambiguated indefinites towards narrow scope

by adding the adjective *different*, e.g., *a different caregiver/child*. Such *different*-marked indefinites are acceptable in discourse-initial (out of the blue) sentences when they take narrow scope relative to distributive quantifiers (among other kinds of licensors, [Carlson 1987](#) among others). These conditions are exemplified in (15) below.

- (15) a. Subject:*different*
Every night, a different caregiver comforted a child.
The caregivers wanted the child to get some rest.
- b. Object:*different*
Every night, a caregiver comforted a different child.
The caregiver wanted the children to get some rest.

In (15a), the indefinite subject is unambiguously treated as having narrow scope with respect to the preceding distributive quantifier while in (15b), the indefinite object must take narrow scope relative to the quantifier.

Predictions

Given previous findings ([Filik et al., 2004](#); [Paterson et al., 2008](#), a.o.), we expect that SUBJECT:PL should lead to an increase in reading times and/or regressions compared to SUBJECT:SG. The same should hold for OBJECT:PL when compared to OBJECT:SG. This difference between plural and singular is expected because: (i) pluralities require inverse scope, (ii) plural nouns are longer, (iii) they are semantically more complex, (iv) there is a number mismatch between the definite and its indefinite antecedent.

Our main focus is on the interaction of SUBJECT:PL and OBJECT:PL. Suppose that it is true that readers use the number information to reanalyze a total and transitive scopal representation that the processor builds. Furthermore, suppose that this reanalysis incurs processing cost. In that case, SUBJECT:PL

and OBJECT:PL should not additively increase reading times and regressions, rather, their combination should have a facilitating effect: SUBJECT:PL and OBJECT:PL should *decrease* reading times and regressions, compared to the sum of the main effects for SUBJECT:PL and OBJECT:PL. The decrease should be visible on the object or words following the object, but it could also be visible in late measures, e.g., re-reading times, on regions preceding the object. The decrease is expected because SUBJECT:PL establishes the inverse scope of the distributive quantifier, so OBJECT:PL, which is compatible with this analysis, causes no further difficulties regarding the interpretation of scope. In terms of the analysis to be discussed, this theory predicts a negative interaction of SUBJECT:PL \times OBJECT:PL. No interaction is expected for underspecification theories of scope, or for theories in which only mental models are reanalyzed, because SUBJECT:PL does not establish scope for the object under these accounts. Also, no interaction is predicted if processing cost has nothing to do with the scope reanalysis (see also Table 1). We'll discuss in more detail later whether a negative interaction could be caused by factors other than scope (re)analysis.

Finally, since late disambiguation of scope should be more taxing for the processor than early disambiguation, we expect that in the second sentence, SUBJECT:PL, OBJECT:SG should increase reading times and/or regressions compared to SUBJECT:*different*. The same should hold for SUBJECT:SG, OBJECT:PL and OBJECT:*different*.

Method

Participants

27 native English speakers participated in the experiment, all of them undergraduate students at UC Santa Cruz. They received course credit for their participation. Each participant had normal or corrected-to-normal vision and

was naive with respect to the purpose of the study.

Materials

The eye-tracking experiment consisted of 7 practice items, 39 experimental items and 67 fillers, for a total of 113 stimuli. Furthermore, there were 33 comprehension questions, controlling that participants paid attention to the stimuli: 12 comprehension questions associated with experimental items and 21 questions associated with fillers. The fillers, like experimental items, consisted of two sentences. Most often, the fillers included simple indefinites or definites, but quantifiers like *every*, *no* and *all* appeared in 25 fillers (an example is given in (16a)), so the difference between the fillers and experimental items would not be signalled by the presence/absence of distributive quantification. 20 fillers were experimental items from a separate experiment testing *swarm*-type verb alternations (an example is given in (16b)). These items were different enough, we believe, to not interfere with the experiment on quantifier scope.

- (16) a. Every park in the area had at least one forest fire last summer. The fires were all pretty small, fortunately.
- b. A field is bursting with green fruits. It is also crawling with bugs.

Each experimental item appeared in six conditions (see, for example, (14) and (15)). Six lists were created out of the stimuli by rotating the items through the conditions across the lists in a standard way, so that: (i) no item appeared more than once in any list, (ii) the six lists exhausted all the conditions of every item, (iii) in every list, six consecutive items cycled through the six conditions. The participants were rotated through the six lists (each participant saw exactly one list). The stimuli in the list were randomly ordered for each participant. We note that the design was not fully balanced since we had 39 items. However, missing data points are not unusual in eye tracking and thus, even a fully bal-

anced design would likely yield an unbalanced result data set. In any case, such a design is hardly an issue for our statistical analysis since we use mixed-effects models, which do not need the balanced design for the proper estimation.

The presented sentences appeared on one or two lines. If an item had to be split between two lines because of its length, we ensured that the line break occurred more than two words after the object of the second sentence (the wrap-up region).

There was a short questionnaire after the eye-tracking study, establishing the acceptability of the items used in the eye-tracking experiment. The questionnaire consisted of 3 practice items, 39 experimental items identical to the items in the eye-tracking experiment and 41 fillers taken from the eye-tracking experiment (83 stimuli in total). Six lists were created out of the stimuli in the same way as in the eye-tracking experiment. Each participant saw only one list, with stimuli randomly ordered.

Procedure

The eye-tracking experiment started with the calibration of every participant. After the calibration, the participant read the practice items. Each practice item appeared as soon as the participant fixated a small black rectangle displayed on the screen at the beginning of the line (the gaze trigger). Every stimulus consisted of two sentences. After the participant read the stimulus, s/he had to press a controller button to move to the yes/no comprehension question or to the next stimulus sentence, whichever was the case. Comprehension questions were answered by pressing one of two controller buttons. After the practice session, the actual experiment began. The setup was identical: as in the practice part, each stimulus in the experiment consisted of a two-sentence discourse, possibly followed by a question. Every stimulus appeared after the successful fixation of the gaze trigger. If participants fixated the gaze trigger and the stimulus

sentence did not appear, the experimenter recalibrated the eye-tracker. The whole experiment lasted approximately 45 minutes. Participants were informed about the length of the experiment beforehand, and they were encouraged to take one or more short breaks during the experiment.

Eye movements were recorded with an EyeLink 1000 machine in desktop mount (chin-rest mode) and monocular recording with a 35 mm lens. The movements of the right eye were recorded at a sampling rate of 1kHz. The calibration was done on a 9-point grid. Stimuli were presented in 15 point non-anti-aliased Courier on a 19-inch Dell monitor with a refresh rate of 85 Hz. All letters in the stimuli were in lower case, except for the first letter in the sentence. The text was black, displayed against a uniform light-gray background.

Immediately after the eye-tracking experiment, 26 out of 27 participants filled in a questionnaire that tested the acceptability of the items presented in the eye-tracking experiment. Each participant had to judge the acceptability of the stimuli on a scale from 1 (worst) to 5 (best). The instructions specified that when rating the sentences, the participants should consider whether they might say / write these sentences or whether they think others might say / write them and that “[t]here are no right or wrong answers beyond your own intuitions”. The questionnaire was completed in the lab on a local computer. We used Alex Drummond’s IBEX platform to implement it (see <http://code.google.com/p/webspr/>). Completing the questionnaire took approximately 10 minutes. Participants took the whole experiment (the eye tracking and the acceptability judgment task) individually.

Results

Comprehension questions and the acceptability judgment task

On average, the participants answered 88% of the comprehension questions correctly and no participant answered less than 74% correctly. We kept all participants for the subsequent analysis.

One item was excluded from all data analyses because it contained a typo. Thus, all subsequent results are based on the remaining 38 experimental items.

The results of the acceptability judgment task that participants filled after the eye-tracking experiment are summarized in Figure 1. We show means and standard errors of the judgments in each condition (assuming for ease of presentation that the response variable was continuous and not ordered categorical, which is customary). Recall that the discourses were judged on 1-5 scale, with 1 the lowest score and 5 the highest. As one can see, the narrow scope of the subject was judged as worse than both early-disambiguation conditions and the ambiguous scope of the indefinite subject.

[Figure 1 about here.]

This was confirmed by the subsequent statistical analysis. Given the ordered categorical nature of the data, we used mixed-effects ordered probit models to statistically analyze the data (using the `ordinal` R package, see [R Core Team \(2013\)](#) and [Christensen \(2013\)](#)). The model had one predictor, `CONDITION`, which was a factor with six levels: `SUBJECT: different`, `OBJECT: different` and the four late-disambiguation conditions of main interest (`SUBJECT: SG or PL` and `OBJECT: SG or PL`), with `SUBJECT: different` as the reference level. The model also included maximal (intercept and slope) random effects for subjects and items. The model revealed a significant effect of `SUBJECT: PL`, which was less acceptable regardless of the status of the object (for `SUBJECT: PL`,

OBJECT:PL, $\beta = -.76$, $SE = 0.34$, $p = .02$, for SUBJECT:PL, OBJECT:SG, $\beta = -.76$, $SE = 0.33$, $p = .02$). This result is compatible with previous findings that it is dispreferred to let a distributive quantifier take inverse scope over the subject (see [Ioup 1975](#), [VanLehn 1978](#), [Gil 1982](#) and [Gillen 1991](#), a.o.). No other effect was significant. In particular, our study did not reveal that the interaction of the narrow scope interpretation of the subject and the object affects acceptability judgments. We now turn to the eye-tracking study, which will provide more subtle measures to study the interaction.

Eye-tracking

We analyzed four reading time measures:

- first-pass reading time: the sum of (the duration of) the fixations in the region before the region is exited (in any direction);
- go-past reading time: the sum of all the fixations since the first fixation in the region until the region is exited to the right, including re-reading earlier parts;
- total reading time: the sum of all fixations in the region;
- re-reading time: total time minus first-pass time.

We also examined the probability of first-pass regression from a given region and the probability of re-reading a region, i.e., entering a region more than once.

Only the second sentence is of interest given our focus on disambiguation. This sentence was split into the regions shown in [Table 2](#) for the purpose of the analysis.

[Table 2 about here.]

Table 3 summarizes the means and standard errors of the reading measures for the regions in the second sentence.

[Table 3 about here.]

Finally, table 4 summarizes the results of our data analysis. In this summary, we set aside the two early-disambiguation conditions (we will return to them in due course) and focus only on the four late-disambiguation conditions.

The summary reports the maximum likelihood estimates (MLEs) and associated p -values of appropriate mixed-effects generalized linear models: mixed-effects linear regression models for the continuous reading-time data (with response variable log-transformed to ‘correct’ for right skewness) and mixed-effects logistic regression models for the categorical ‘probability of regression’ and ‘probability of re-reading’ data. All models included the full fixed-effect structure: the main effect of subject interpretation (SUBJECT: PL vs. SUBJECT: SG, SG being the reference level), the main effect of object interpretation (OBJECT: PL vs. OBJECT: SG, SG being the reference level), and their interaction. They also included the maximal subject and item random-effect structure for which the estimation procedure converged. We estimated all models with the `lme4` R package (see Bates et al. (2013)). The `lme4` package does not provide degrees of freedom for t -values of mixed-effects linear regression models based on which we can obtain approximate p -values (rightfully so), and also does not provide a Markov Chain Monte Carlo (MCMC) sampling procedure to calculate approximate p -values for the fixed effects in models with both intercept and slope random effects. To obtain more specific information about the level of significance of studied effects, we repeated the data analysis in a Bayesian framework and calculated ‘ p -values’ based on samples from the joint posterior distribution of the fixed and random effects of our hierarchical models. The estimates of the joint posterior distributions for all our models were obtained with JAGS (see

Plummer (2013) for further details). A typical Bayesian hierarchical model that we used to estimate these ‘ p -values’ is provided in the supplementary files; note in particular the vague / low information priors we used for both the fixed and the random effects. It is worth noticing that ‘ p -values’ obtained this way line up closely with t -values gathered from `lme4` models, as one would hope to be the case (e.g., all the effects with $t > 2$ are significant, the effects with t -values at 1.9 or lower are not). Only ‘ p -values’ smaller than .1 are reported. Significant effects at the conventional $\alpha = .05$ level are boldfaced.

[Table 4 about here.]

Given previous findings, we expected higher reading measures in the plural subject condition than in the singular subject condition. Such a significant increase was, at its earliest, observed in total times and the probability of re-reading in the Subject and Verb regions. Importantly, the increase in the probability of re-reading was only present when the object was singular. When the object appeared in plural, there was no increase, as witnessed by the negative interaction of SUBJECT:PL and OBJECT:PL that had almost the same magnitude as the main effect SUBJECT:PL ($\beta = -0.8$ and $\beta = 0.7$, respectively). To show this more perspicuously, we also graphed all measures/regions with significant interactions in Figure 2. The graph of the probability of re-reading for Subject and Verb clearly shows that SUBJECT:PL only incurs cost when the object is singular.

[Figure 2 about here.]

We also expected higher reading measures in the plural object condition than in the singular object condition. These were first found on the object itself, where OBJECT:PL significantly increased total times, the probability of regression and the probability of re-reading. In the same region, the probability

of regression and the probability of re-reading also showed a spillover effect of SUBJECT:PL. Yet again, we observe a negative interaction on these measures, showing that the two-plural condition is not additive here. In fact, the two-plural condition does not exhibit more regressions or re-readings than the conditions with only one plural argument (see the measures on Object in Figure 2). One possible interpretation is that the regressions and re-readings in the two-plural condition show the spillover cost of the plural subject, and the plural object incurs no extra difficulties.

The cost of SUBJECT:PL did not spill over to any reading time measures beyond the object to show a significant effect. In contrast, the cost of OBJECT:PL was also detectable in significantly longer first-pass and total times for the Spillover region. Like several previous measures, first pass also revealed a negative interaction: the cost of plural objects disappeared when subjects appeared in plural. Thus, first pass of the Spillover region paints the same picture as the regressions and re-readings on the object, the only difference being that the spillover cost of SUBJECT:PL itself is diminished and non-significant.

Finally, the spillover cost of plural subjects was revealed by a significant increase in the probability of regression and re-reading in the Wrap-up region. Yet again, both measures showed a significant negative interaction of plural arguments. That is, even in Wrap-up region, the two-plural condition did not have more regressions and re-readings than the conditions with one plural (see Wrap-up in Figure 2).

Because narrow scope is forced by plural morphology, it is not clear whether the increased time and the increased regressions were caused by the computation of inverse scope or due to differences between singular and plural morphology (differing length, frequency, number of referents etc.). We therefore did two pairwise comparisons involving our two early-disambiguation conditions: (i) a

comparison between SUBJECT:PL, OBJECT:SG and SUBJECT:*different*, and (ii) one between SUBJECT:SG, OBJECT:PL and OBJECT:*different*. The models had a single fixed effect, DISAMBIGUATION (EARLY vs. LATE, the former being the reference level) and intercept and slope random effects for subjects and items. Even though the late-disambiguation conditions were numerically slower than their corresponding early-disambiguation conditions in several reading measures in Subject and Verb and Object (see Table 3), the pairwise comparison did not reach significance in any region. Thus, we lack conclusive independent evidence indicating that the processing difficulties associated with SUBJECT:PL or OBJECT:PL are solely due to late disambiguation rather than to any plural-morphology related effects. We return to this issue in the discussion subsection.

Discussion

The analysis reveals that in several regions and measures, the cost of plural is not additive: the condition with two plurals is at most as costly as the conditions with a single plural argument.

Why do we see this negative interaction? It cannot be a consequence of better acceptability of the narrow scope of both arguments in our items, given that in our acceptability study, no difference between SUBJECT:PL, OBJECT:PL and SUBJECT:PL, OBJECT:SG was found: the two conditions were judged as equally good. Instead, an explanation in terms of processing difficulties, discussed previously, suggests itself here: assuming that the scope hierarchy is transitive and total (see (8)), the narrow scope object becomes the default interpretation when the subject is interpreted as having narrow scope; it does not matter how many arguments are plural, the inverse-scope reanalysis takes place only once. This explains why the combination of SUBJECT:PL and OBJECT:PL is not more costly (on the relevant measures) than the conditions with only one

plural argument. Note also that this explanation implies that the reanalysis of scope is detectable in eye tracking and it can be distinguished from the inherent cost of pluralities, at least in some measures/regions.

An alternative explanation to our findings is that all instances of the negative interaction SUBJECT:PL×OBJECT:PL have nothing to do with scope disambiguation. One possibility is that pluralities are slower/harder to process but the plural morphology on the second argument is primed by the plural morphology on the first argument and this priming gives rise to the observed pattern.

If we want to maintain that our results reflect deeper properties of scope, we have to exclude the possibility that the negative interaction is simply due to the use of two plural arguments.

This issue is even more pressing when we consider the two pairwise comparisons between the early-disambiguation conditions and the corresponding late-disambiguation conditions. As we already noted, the timing of the disambiguation (early vs. late) had no significant effect on reading measures in the second sentence. This could be taken as evidence that scope disambiguation in the second sentence causes no processing difficulties, which would yet again point to the conclusion that the negative interaction between SUBJECT:PL and OBJECT:PL is not due to their scope interpretation, but to other factors.

There is however another difference between our early vs. late disambiguation conditions. For all the late-disambiguation conditions, e.g., (17a) below, the first sentence and the second sentence had a parallel structure: each of them began with the subject, followed by the verb, the object and the quantifier in adjunct position. The early-disambiguation conditions, however, placed the distributive quantifier in the sentence-initial (topical) position, e.g., (17b,17c), thus breaking the parallelism between the first and the second sentence. Since par-

allelism speeds up reading (Frazier et al., 1984, 2000), its presence in (17a) and its absence in (17b,17c) might obscure any scope-related effect.

- (17) a. A caregiver comforted a child every night.
The caregiver(s) wanted the child(ren) to get some rest.
- b. Subject:*different*
Every night, a different caregiver comforted a child.
The caregivers wanted the child to get some rest.
- c. Object:*different*
Every night, a caregiver comforted a different child.
The caregiver wanted the children to get some rest.

Thus, Experiment 1 establishes that SUBJECT: PL facilitates the processing of OBJECT: PL, but it leaves two issues unresolved. First, it cannot determine whether the facilitation is due to scope processing or is caused by ‘shallower’ factors tied to the repeated use of plural morphology. Second, it lacks any independent evidence that late disambiguation, as compared to early scope disambiguation, is taxing for human processor. Experiment 2 addresses these issues.

Experiment 2: A self-paced reading study

Introduction

Experiment 2 was conducted to check if the results of Experiment 1 can be replicated using a self-paced reading task (Just et al., 1982) and to address the two outstanding issues discussed above.

In contrast to Experiment 1, Experiment 2 introduces a new experimental manipulation, the presence or absence of CONTEXT. The items in which the

context was present were identical to the items in Experiment 1. They consisted of two sentences: the first sentence included two singular indefinites and a distributive universal quantifier; the second sentence disambiguated the scope of the indefinites using plural arguments. Furthermore, in the case of early disambiguation, the items included *different* to disambiguate scope in the first sentence just as in Experiment 1. However, to ensure the parallelism between the first and the second sentence, distributive quantifiers were not preposed (as in Experiment 1) but were left in sentence-final (non-topical) position. An example item is provided in (18) and (19).

(18) CONTEXT: YES

A caregiver comforted a child every night.

a. SUBJECT: PL, OBJECT: PL:

The caregivers[pl] wanted the children[pl] to get some rest.

b. SUBJECT: PL, OBJECT: SG:

The caregivers[pl] wanted the child[sg] to get some rest.

c. SUBJECT: SG, OBJECT: PL:

The caregiver[sg] wanted the children[pl] to get some rest.

d. SUBJECT: SG, OBJECT: SG:

The caregiver[sg] wanted the child[sg] to get some rest.

(19) a. SUBJECT: *different*:

A **different** caregiver comforted a child every night.

The caregivers[pl] wanted the child[sg] to get some rest.

b. OBJECT: *different*:

A caregiver comforted a **different** child every night.

The caregiver[sg] wanted the children[pl] to get some rest.

The last two conditions (19a,19b) force disambiguation in the first sentence

due to the presence of *different*. As noted before, *different* normally requires an antecedent in the previous discourse to be acceptable, but it can also appear without an antecedent if it is in the scope of a distributive quantifier. Since no discourse antecedent is available for *different* in (19a,19b) but a distributive quantifier is present in these sentences, the only felicitous option is to interpret the distributive quantifier *every night* as having scope over *different*. Thus, this experimental setup preserves the contrast between early and late disambiguation, but eliminates orthogonal differences between the early-disambiguation and the late-disambiguation conditions.

The stimuli in the CONTEXT:NO conditions included only the second sentences of the corresponding stimuli in the CONTEXT:YES conditions, as shown in (20) below.

(20) CONTEXT:NO

a. SUBJECT:PL, OBJECT:PL:

The caregivers[pl] wanted the children[pl] to get some rest.

b. SUBJECT:PL, OBJECT:SG:

The caregivers[pl] wanted the child[sg] to get some rest.

c. SUBJECT:SG, OBJECT:PL:

The caregiver[sg] wanted the children[pl] to get some rest.

d. SUBJECT:SG, OBJECT:SG:

The caregiver[sg] wanted the child[sg] to get some rest.

Predictions

Since only the second sentences from Experiment 1 were included in CONTEXT:NO, the plural arguments did not perform any scope disambiguation. The comparison between the conditions with and without the context therefore reveals whether the facilitation of SUBJECT:PL on the processing of OBJECT:PL

is due to scope disambiguation or just due to the repetition of plural arguments. If the latter is the case, we should observe a negative interaction of SUBJECT:PL and OBJECT:PL regardless of the presence or absence of the context. If SUBJECT:PL facilitates OBJECT:PL only when it disambiguates scope, the negative interaction should take place when the context is present and should be missing when the context is absent.

Method

Participants

The participants were 88 native English speakers (44 participants for CONTEXT:NO, 44 participants for CONTEXT:YES), all of them undergraduate students at UC Santa Cruz. They received course credit for their participation.

Materials

The CONTEXT:NO and CONTEXT:YES conditions were run in a between-subject design because of the different structure of the stimuli: they consisted of one-sentence vs. two-sentence discourses, respectively. This manipulation would have been too prominent in a within-subject experimental design.

In both versions of the experiment, there were 4 practice items and 39 experimental items. The experimental items in CONTEXT:YES differed from the items used in Experiment 1 only with respect to the *different* conditions, as indicated above. We also corrected the typo from Experiment 1, so all 39 items could be analyzed. The items in the CONTEXT:NO conditions were identical to the second sentences of the items in CONTEXT:YES. However, the verbs in four items in CONTEXT:YES were in past perfect, which was grammatically adequate in the given context but were not felicitous out of the blue, so the verbs in these items were changed to simple past in the CONTEXT:NO conditions.

Furthermore, the experiment included 67 fillers in the version with context and 65 fillers in the other version. The fillers in the first version were taken from Experiment 1. In the version with no context, 47 fillers were first sentences of fillers from Experiment 1. The remaining 18 fillers were items of a separate experiment, testing the inverse-scope capabilities of quantifiers *each* and *every*. Since the version with no context did not test scope interpretation, such fillers should not affect the reading of the experimental items.

There were 32 comprehension questions in each version of the experiment, 12 of which targeted experimental items.

Six lists of stimuli were created for the version with context, and four list for the version without context. Every item occurred in each list exactly once, and the lists were obtained by rotating the conditions in which the items occurred. Each participant saw exactly one list. The order of presentation of the stimuli in the list was randomized for each participant.

Procedure

The self-paced reading experiment was run on a local (UCSC hosted) installation of the IBEX platform. When a stimulus first appeared on the screen, all the word characters were masked by a series of dashes. When participants pressed the space bar once, the first word of the stimulus was unmasked, i.e., the first group of dashes was replaced by the first word of the stimulus. When participants pressed the space bar a second time, the first word was changed back to dashes and the second word was revealed; and so on. Upon reaching the end of the stimulus, either a comprehension question was displayed or the next stimulus. All participants received a link to the experiment and completed the self-paced reading task on-line. The task lasted no more than 30 minutes.

Results and discussion

Three participants answered 75% of comprehension questions or less correctly, and we excluded them from the subsequent analysis. We also excluded four other participants whose mean (log-transformed) reading times were more than 2 standard deviations away from the grand mean for all subjects. The final number of participants considered in the analysis was 81.

Just as in Experiment 1, the response variable (reading time in ms.) is continuous, so we use mixed-effect models to analyze the data. We have main effects for SUBJECT NUMBER, OBJECT NUMBER (both with SG and PL as possible values, and SG as the reference level) and CONTEXT (with CONTEXT:NO as the reference level), and the two-way and three-way interactions of the main effects. The models also included the maximal random-effect structure for subjects and items for which the estimation procedure converged.¹

The dependent variable was residualized log reading times (RTs). The residualized log RTs were obtained by estimating a mixed-effects linear model with log RTs as the response variable, word length (in characters) and word position as fixed effects, and intercept random effects for subjects (see Trueswell et al. 1994 for motivation). Data from both the experimental items and the fillers was used in the estimation of the residualized log RTs for a greater precision.

The words that are of main interest in the current study are boldfaced in (21a) and (21b) below (for both kinds of contexts). For completeness, however, we will also show measures for words preceding the object. However, we do

¹CONTEXT: YES and its interaction with other parameters could not be a slope coefficient for subjects because its presence/absence was a between-subject manipulation. But the models had to be further simplified due to issues with convergence. The maximal model used which converged for most measures included the random intercept and the SUBJECT: PL and OBJECT: PL random slopes for subjects, and the random intercept and the SUBJECT: PL, OBJECT: PL, CONTEXT: YES and OBJECT: PL×CONTEXT: YES random slopes for items.

not discuss the words after *get* since several items had a short spillover region consisting of only two words after the object. Because of this, statistical power in the regions following *get* is diminished and no effect is significant.

- (21) a. CONTEXT: YES
A caregiver comforted a child every night. The caregiver(s) wanted the **child(ren) to get** some rest.
- b. CONTEXT: NO
The caregiver(s) wanted the **child(ren) to get** some rest.

We also compared the early-disambiguation conditions SUBJECT:*different* and OBJECT:*different* with their late-disambiguation counterparts. We will discuss these pairwise comparisons separately.

The means and standard errors are graphically summarized in Figures 3 and 4.

[Figure 3 about here.]

[Figure 4 about here.]

The MLEs of the slope coefficients in the corresponding mixed-effects models are provided in Table 5.² Table 5 shows that there is only one significant main effect, that of CONTEXT, whose presence leads to a speed-up (as seen from the negative slope of CONTEXT: YES). This effect, albeit interesting, is orthogonal to our inquiry. The facilitating role of the context can be explained by the fact that the first sentence allows readers to more easily predict upcoming words in the second sentence.

[Table 5 about here.]

²We do not include the effect of OBJECT NUMBER in the analysis for the words preceding the object, since that experimental manipulation takes place only on the object itself.

Aside from this effect, we also see several interactions on the second word after the object. When the context is absent, there is a positive borderline-significant interaction of SUBJECT:PL and OBJECT:PL. The crucial fact is that the interaction is not negative, which shows that a plural subject in itself does *not* facilitate the processing of the plural object. If anything, the marginally significant result suggests that the plural subject makes the processing of the plural object harder when the arguments serve no scope-disambiguating role.

In contrast, we observe a three-way *negative* interaction of the plural subject, plural object and CONTEXT:YES. That is, the plural subject facilitates the processing of the plural object, but only when plural arguments are used to disambiguate scope. This finding strengthens the conclusion we have drawn on the basis of Experiment 1. We both replicate the Experiment 1 results (in the CONTEXT:YES condition) and, given the findings in the CONTEXT:NO condition, we can conclude that the facilitation should be attributed to scope disambiguation and not to the repeated occurrence of plural morphology.

Finally, the same word also shows a positive interaction of plural objects and context. This slowdown is likely the reflection of inverse-scope computation since the plural object in CONTEXT: YES is the first signal of inverse scope when the subject is singular. We thus see that the plural object indeed is taxing for the processor, but only if it disambiguates scope and the subject did not already force the inverse-scope computation.

We turn now to the two pairwise comparisons in CONTEXT: YES between the early-disambiguation conditions and the corresponding late-disambiguation conditions: (i) SUBJECT:SG, OBJECT:PL vs. OBJECT:*different* and (ii) SUBJECT:PL, OBJECT:SG vs. SUBJECT:*different*.

We start with the first comparison, i.e., between (22a) and (22b). We are mainly interested in reading times for the boldfaced words but for completeness,

we compare reading times for the entire second sentence – see Figure 5.

- (22) a. SUBJECT:SG, OBJECT:PL:
A caregiver comforted a child every night.
The caregiver wanted the **children to get** some rest.
- b. OBJECT:*different*:
A caregiver comforted a different child every night.
The caregiver wanted the **children to get** some rest.

We expect that the early scope disambiguation caused by *different* should lead to faster reading times than the late disambiguation. The difference should be observable on the second word after the object, which showed the effect of the inverse-scope object disambiguation. This prediction is borne out. The model with one fixed effect DISAMBIGUATION (reference level: EARLY), and random intercepts and slopes for subjects and items reveals a significant effect of LATE disambiguation, ($\beta = 0.07, p = .03$), on the second word after the object.

[Figure 5 about here.]

In contrast, the second comparison shows results that seem somewhat surprising at first. Recall that this case compares the early and late disambiguation of the scope of the subject quantifier. Thus, we are interested in the reading times for the subject of the second sentence and the following words:

- (23) a. SUBJECT:PL, OBJECT:SG:
A caregiver comforted a child every night.
The **caregivers wanted the child to get** some rest.
- b. SUBJECT:*different*:
A different caregiver comforted a child every night.
The **caregivers wanted the child to get** some rest.

SUBJECT:*different* is numerically *slower* on the subject noun and the following word and the effect is borderline significant. The model with one fixed effect DISAMBIGUATION (reference level: EARLY) and intercept-only random effects for subjects and items³ has a borderline-significant LATE slope coefficient, $\beta = -0.04, p = .07$ for both words; see also Figure 6.

[Figure 6 about here.]

This weak result might seem surprising at first, but it is in fact expected. The crucial observation is that the early and late scope disambiguation for subjects differ only slightly in timing. In the former case, the disambiguation towards inverse scope can happen at the earliest when the word *night* (the distributive-quantifier noun) is read.⁴ In the latter case, the disambiguation can happen at the earliest when the word *caregivers* (the subject noun) is read. The distance between the two disambiguation points is only two words. We noted above that the effect of the inverse-scope computation on the object was visible with a delay of two words, and such a delay makes it very hard to detect any effects in this case.

³Slope random effects were estimated at 0 or extremely close to 0 and thus, the effects were negligible.

⁴An anonymous reviewer points out that readers might anticipate distributive quantifiers and the necessity of their inverse scope already when they encounter *different*. It is unclear to us whether readers really have such expectations since there are other means to satisfy the requirement of *different* aside from assigning a distributive quantifier higher scope. For example, *than*-clauses can provide the necessary licensing for *different*, as in the following example:

- (1) A different set of questions should be addressed than those that have occupied Fukayama's respondents.

Even if readers did anticipate the inverse scope of a quantifier, they cannot compute the correct inverse scope (i.e., the correct interpretation) until they encounter the actual quantifier in the text.

Given the possibility of spillover from the end of the first sentence, SUBJECT: *different* should be associated with higher reading times on the subject itself, which is borne out, albeit the effect is only borderline-significant. We also see that two words after the subject, the reading times become more similar: there is no borderline significant effect any more ($p > .1$); see also Figure 6. On the last word (*get*), the difference between the two conditions numerically increases even though the corresponding p -value remains greater than .1.

In conclusion, the results of Experiment 2 replicate the facilitation observed in Experiment 1. They furthermore provide novel evidence that the facilitation is a consequence of scope disambiguation since the facilitation is only observed when plural arguments disambiguate scope. Finally, Experiment 2 shows that the early disambiguation towards object narrow scope speeds up reading times as compared to its late scope disambiguation. This supports the conclusion that the processing cost associated with the plural object in CONTEXT: YES is due to its scope disambiguating role.

General discussion

Experiments 1 and 2 establish that the inverse-scope interpretations of subjects and objects are related: if the subject forces inverse scope with respect to an adjoined distributive quantifier, the inverse-scope interpretation of the object with respect to the same quantifier stops being costly. We argued that this requires a model in which the inverse-scope interpretation of the subject creates the inverse-scope interpretation of the object as a by-product. This is true for hierarchical scope representations irrespective of whether they are based on linear order (Johnson-Laird, 1969, a.o.), the thematic-role hierarchy (Jackendoff, 1972, a.o.), or the grammatical-role hierarchy (Reinhart, 1983, a.o.).

In contrast to that, underspecification theories of scope (see, for exam-

ple, [Ebert 2005](#) and references therein) do not postulate any default hierarchy. Therefore, the narrow scope of the subject with respect to an adjunct does not affect the scope interpretation of the object in any way. Consequently, underspecification theories of scope don't explain the data discussed in this paper.⁵

The same problem arises if the inverse scope interpretation would not result in the change of the scope hierarchy but in the direct change of the interpretation / mental model. Since the models in [Johnson-Laird et al. \(1989\)](#) or the ones often used in formal semantics do not store any dependency between subjects and objects, there is no way to ensure that the narrow-scope interpretation of the subject should influence the scope of the object.

Aside from sharpening grammatical theories of quantifier scope, our results bear directly on two issues relevant for the processing of scope. First, they provide evidence that the reanalysis of the default / surface scopal representation towards inverse scope must be rapid, and the new scopal information is quickly integrated. The narrow-scope reanalysis of the subject, required by the plural *caregivers* in (24) in our experiments, must be achieved by the time the object is read, since this is the point at which we already observe its facilitation effects in Experiment 1.

(24) The caregivers wanted the children to get some rest.

Second, our results provide an answer to the question of what causes processing difficulties when a plural definite signals an inverse-scope interpretation. Finding such an answer is important since no consensus has been reached with respect to this issue. We noted that the processing difficulties often observed in plural disambiguations towards inverse scope ([Kurtzman and MacDonald, 1993](#);

⁵Another argument against underspecification theories of quantifier scope appears in [Radó and Bott \(2012\)](#) based on a less conventional experimental methodology (incremental truth-value judgment task).

Tunstall, 1998; Anderson, 2004, a.o.) can be tied to inverse-scope computation or, as Filik et al. (2004) note, they might be caused by a mismatch in number between the plural definite and its singular antecedent.

It is hard, if not impossible, to decide between these two possible explanations when studying only sentences with two quantifiers. The only investigation directly targeting this issue that we are aware of involved sentences like (25) below (from Anderson 2004).

(25) Every historian examined a document. The document/the documents...

In these sentences, the surface order of the distributive quantifier and the indefinite is the inverse of the order we considered in the present study. Given this order, the distributive quantifier should have the indefinite in its scope by default. Consequently, the plural continuation *the documents* should not cause processing difficulties compared to *the document* if it was inverse-scope computation what was taxing, but such difficulties should still be observed if they were caused by morphological mismatch.

Unfortunately, the results reported in the previous literature that bear on this issue are not concordant. Tunstall (1998) finds slower reading times for plural continuations relative to singular continuations in the *a*»*every* order, and no difference in the *every*»*a* order, which follows if the processing cost is caused by inverse-scope computation.

In contrast, Anderson (2004) finds that the singular continuation is read slower than the plural continuation in (25), which is surprising for both theories: the scope computation does not predict this contrast because both continuations are compatible with the surface scope, and the cost due to a morphological mismatch expects the opposite pattern. That is, the results of Anderson (2004) are problematic for the morphological mismatch account and orthogonal to the scope computation account. They signal that another source of semantic pro-

cessing effects might be involved in (25), probably tied to the fact that people disprefer ‘accidental co-reference’.

Finally, Filik et al. (2004) and Paterson et al. (2008) find that regardless of which order is used ($a \gg \textit{every}$ or $\textit{every} \gg a$), the plural continuation causes (comparable) processing difficulties (that is, no interaction between the type of order and type of continuation was detected). This result supports the explanation in terms of morphological mismatch.

Turning to sentences with three quantifiers is important from this perspective, since the data provide novel evidence for the cost of inverse-scope computation and against the morphological mismatch account. More generally, our results argue against other attempts of explaining away inverse-scope processing cost as reflecting nothing else than the greater complexity of plurals relative to singulars (be it in terms of length, reference, morphological makeup etc.): (i) we saw that the combination of two plurals is easier, not more difficult, than the combination of one plural and one singular argument, and (ii) the negative interaction was only observed in Experiment 2 when the plural argument played a scope disambiguating role. Time will tell if this evidence for inverse scope cost is more robust and easier to replicate than the contrast between the $a \gg \textit{every}$ and $\textit{every} \gg a$ orders.

Conclusion

We discussed two experiments that investigated the interpretation of sentences with three quantifiers: two singular indefinites in subject and object positions, and a universal distributive quantifier in adjunct position. Studying the processing of such sentences in real time enabled us to distinguish between different theories of quantifier scope. In particular, it provided evidence for theories in which hierarchical, totally ordered scopal representations are involved in resolv-

ing scope ambiguities. Theories of quantifier scope that lack such hierarchical representations, e.g., underspecification theories of scope, are not consistent with the observed data. Furthermore, theories assuming that the scope disambiguating information is ignored by the processor, or integrated late, or used to update mental models directly, also fail to account for the experimental results. What emerges is a model of quantifier scope processing in which the processor builds, and quickly updates fully specified representations of quantifier scope; crucially, these scopal representations are maintained, updated and passed on both within and across sentence boundaries.

Acknowledgments We would like to thank Nathan Arnett, Donka Farkas, Bart Geurts, Lotte Hogeweg, Joseph King, Arnout Koornneef, Megan Moodie, Milica Petranovic, Floris Roelofsen, Kristen Syrett, Matt Wagers, the audiences of the Amsterdam Colloquium 2013, the Nijmegen Semantics & Pragmatics colloquium, the Cross-linguistic Semantics meeting in Amsterdam, the Workshop on Quantifier Domain Restriction and Epistemic Modals at Rutgers and the participants of the joint *Language, Logic and Cognition, Semantics and Syntax* lab meetings and the S-circle meetings at UC Santa Cruz. Jakub Dotlačil was supported by a Rubicon grant and by the NWO Veni Grant 275.80.005 from the Netherlands Organization for Scientific Research. Adrian Brasoveanu was supported by an SRG grant from the UC Santa Cruz Committee on Research. The usual disclaimers apply.

References

AnderBois, S., A. Brasoveanu, and R. Henderson (2012). The pragmatics of quantifier scope: A corpus study. In A. Aguilar-Guevara, A. Chernilovskaya,

- and R. Nouwen (Eds.), *Proceedings of Sinn und Bedeutung 16*, Cambridge, MA, pp. 15–28. MIT Working Papers in Linguistics.
- Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity*, (Unpublished doctoral dissertation). Northwestern University, Evanston, Illinois.
- Bates, D., M. Maechler, B. Bolker, and S. Walker (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-5.
- Bos, J. (1995). Predicate logic unplugged. In *Proceedings of the 10th Amsterdam Colloquium*, Amsterdam.
- Carlson, G. (1987). Same and Different: some consequences for syntax and semantics. *Linguistics and Philosophy* 10, 531–565.
- Christensen, R. H. B. (2013). ordinal—regression models for ordinal data. R package version 2013.9-30.
- Dwivedi, V. D., N. A. Phillips, S. Einagel, and S. R. Baum (2010). The neural underpinnings of semantic ambiguity and anaphora. *Brain Research* 1311, 93–109.
- Ebert, C. (2005). *Formal investigations of underspecified representations*, (Unpublished doctoral dissertation). King’s College, London.
- Filik, R., K. B. Paterson, and S. P. Liversedge (2004). Processing doubly quantified sentences: Evidence from eye movements. *Psychonomic Bulletin & Review* 11(5), 953–959.
- Fodor, J. D. (1982). The mental representation of quantifiers. In S. Peters and E. Saarinen (Eds.), *Processes, Beliefs and Questions*, pp. 129–164. Dordrecht: Reidel.

- Frazier, L., A. Munn, and C. C. Jr. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research* 29, 343–371.
- Frazier, L., L. Taft, C. Clifton, T. Roeper, and K. Ehrlich (1984). Parallel structure: A source of facilitation in sentence comprehension. *Memory and Cognition* 12, 421–430.
- Frege, G. (1879). Begriffsschrift, eine der arithmetischen nachgebildete formelsprache des reinen denkens. In J. vanHeijenoort (Ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879/1931*. Cambridge, Massachusetts: Harvard University Press, 1967. Translated as Concept Script, a formal language of pure thought modelled upon that of arithmetic, by S. Bauer-Mengelberg.
- Gil, D. (1982). Quantifier scope, linguistic variation, and natural language semantics. *Linguistics and Philosophy* 5(4), 421–472.
- Gillen, K. (1991). *The comprehension of doubly quantified sentences*, (Unpublished doctoral dissertation). University of Durham.
- Heim, I. and A. Kratzer (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.
- Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball (Ed.), *Syntax and Semantics 4*, pp. 37–58. New York: Academic Press.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, Massachusetts: MIT Press.
- Johnson-Laird, P. N. (1969). On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology* 21, 1–13.

- Johnson-Laird, P. N., R. M. J. Byrne, and P. Tabossi (1989). Reasoning by model: The case of multiple quantification. *Psychological Review* 96, 658–673.
- Just, M. A., P. A. Carpenter, and J. D. Woolley (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2), 228–238.
- Kemtes, K. A. and S. Kemper (1999). Aging and resolution of quantifier scope effects. *Journal of Gerontology* 54, 350–360.
- Kurtzman, H. S. and M. C. MacDonald (1993). Resolution of quantifier scope ambiguities. *Cognition* 48, 243–279.
- May, R. (1985). *Logical Form: Its Structure and Derivation*. Cambridge, Massachusetts: MIT Press.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In P. S. Jaakko Hintikka, Julius Moravcsik (Ed.), *Approaches to Natural Language*, pp. 221–242. Dordrecht: Reidel.
- Paterson, K. B., R. Filik, and S. P. Liversedge (2008). Competition during the processing of quantifier scope ambiguities: evidence from eye movements during reading. *The Quarterly Journal of Experimental Psychology* 61, 459–473.
- Plummer, M. (2013). *rjags: Bayesian graphical models using MCMC*. R package version 3-10.
- Pylkkänen, L. and B. McElree (2006). The syntax-semantic interface: On-line composition of sentence meaning. In *Handbook of Psycholinguistics*, pp. 537–577. New York: Elsevier.

- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Radó, J. and O. Bott (2012). Underspecified representations of quantifier scope? In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, and M. Westera (Eds.), *Logic, Language and Meaning: 18th Amsterdam Colloquium*. The Netherlands: Springer.
- Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. Chicago: University of Chicago Press.
- Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics* 10, 123–179.
- Trueswell, J., M. Tanenhaus, and S. Garnsey (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33, 285–318.
- Tunstall, S. (1998). *The interpretation of quantifiers: Semantics and processing*, (Unpublished doctoral dissertation). University of Massachusetts, Amherst.
- VanLehn, K. A. (1978). Determining the scope of English quantifiers. Technical Report.

List of Figures

1	Acceptability task: Means & SEs	45
2	Means and SEs of measures showing a significant negative interaction of SUBJECT:PL and OBJECT:PL	46
3	Self-paced reading, CONTEXT:YES: Means and SEs	47
4	Self-paced reading, CONTEXT:NO: Means and SEs	48
5	Self-paced reading, Early/late disambiguation for objects: Means and SEs	49
6	Self-paced reading, Early/late disambiguation for subjects: Means and SEs	50

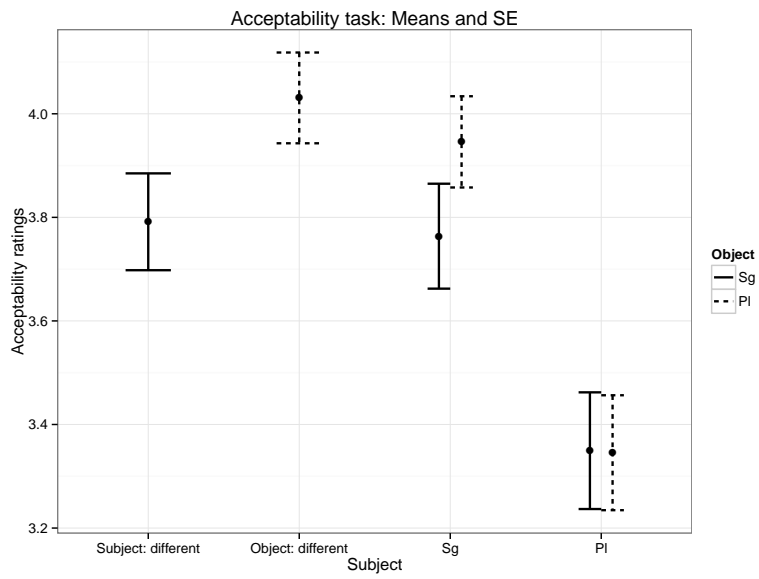


Figure 1: Acceptability task: Means & SEs

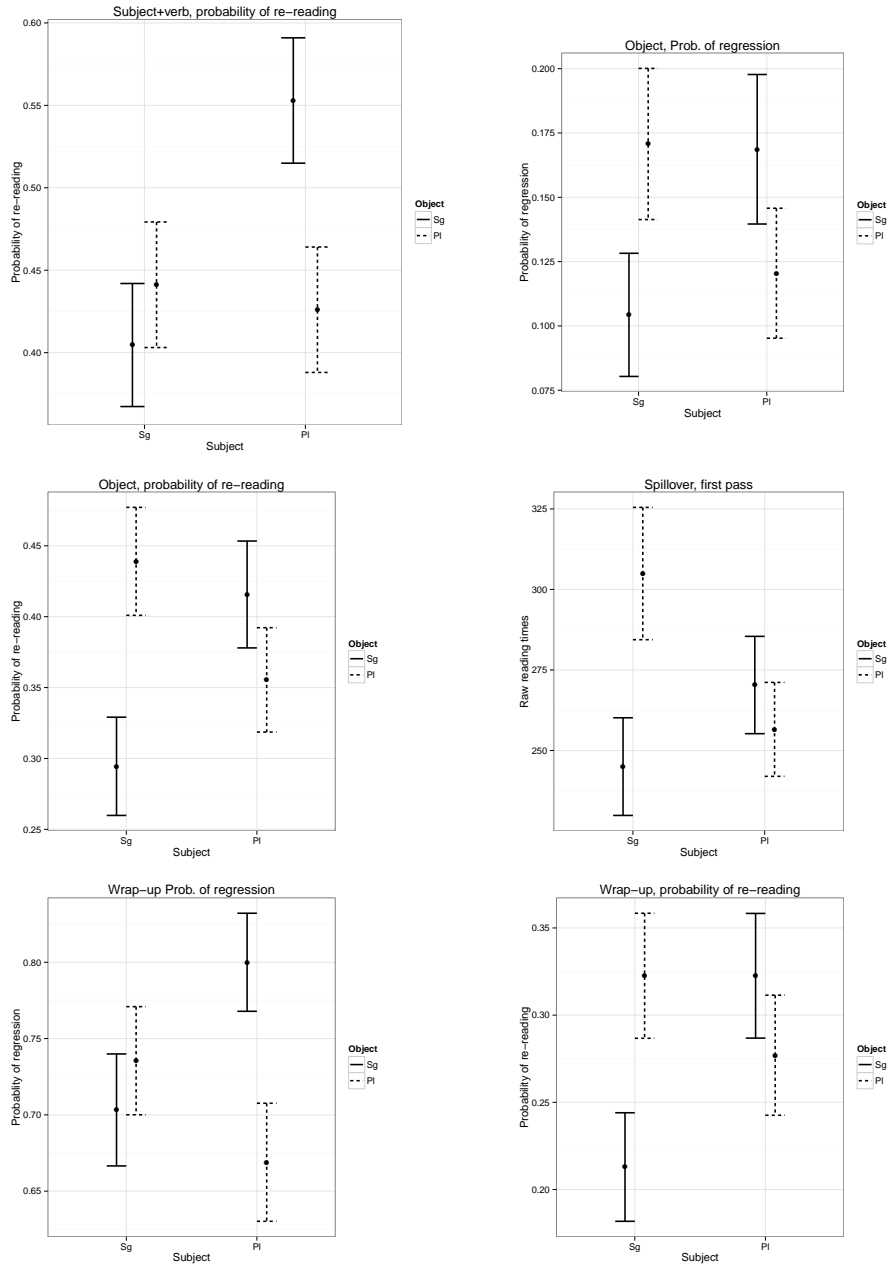


Figure 2: Means and SEs of measures showing a significant negative interaction of SUBJECT:PL and OBJECT:PL

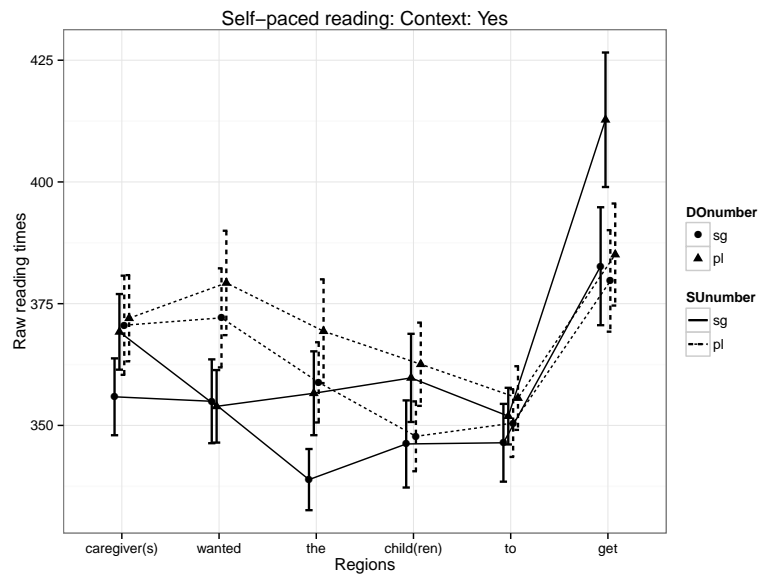


Figure 3: Self-paced reading, CONTEXT:YES: Means and SEs

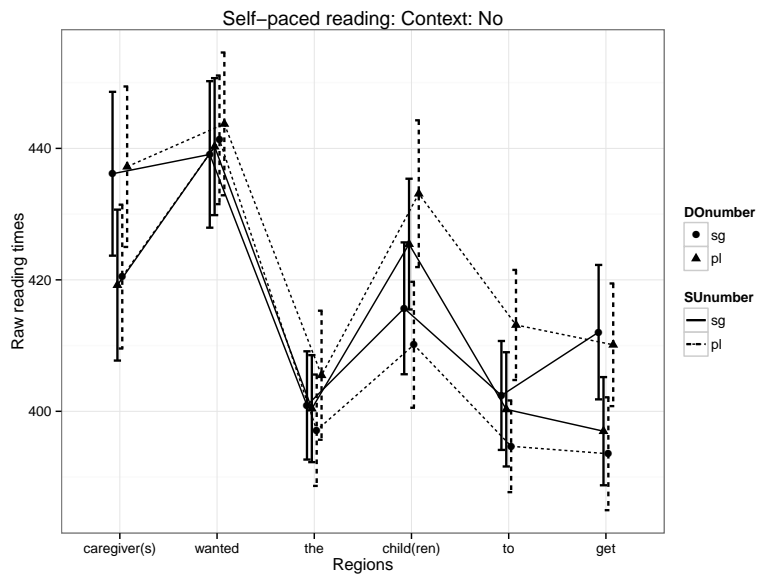


Figure 4: Self-paced reading, CONTEXT:NO: Means and SEs

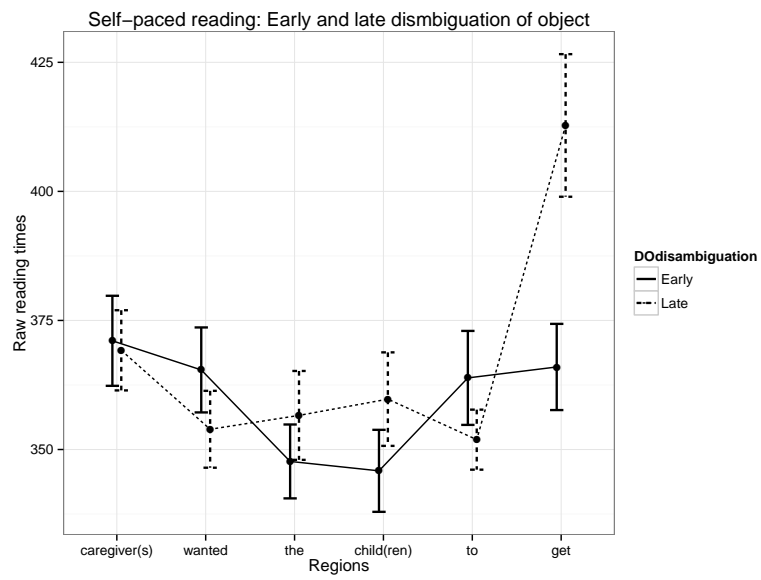


Figure 5: Self-paced reading, Early/late disambiguation for objects: Means and SEs

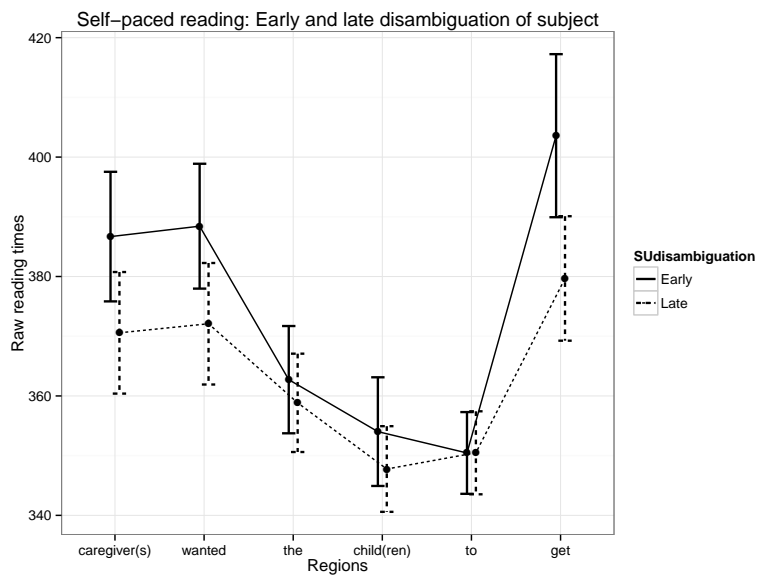


Figure 6: Self-paced reading, Early/late disambiguation for subjects: Means and SEs

List of Tables

1	Predictions of how disambiguating information in (11) affects the continuation in (12)	52
2	Regions in the eye-tracking experiment	53
3	Means & SEs (in ms) in the second sentence	54
4	Slopes and values of significance testing in the second sentence	55
5	Slopes and values of significance testing of the effects	56

	SHOULD NARROW-SCOPE SUBJECT (E.G., 11A) FACILITATE NARROW-SCOPE OBJECT (E.G., 12A)?	
	YES	NO
Theories of scope	Unambiguous scope hierarchy	Underspecified scope (Reyle, 1993; Bos, 1995; Ebert, 2005)
Processing cost	Due to scope reanalysis	Only due to morphological mismatch or other intrinsic features of plurals (Filik et al., 2004)
What is reanalyzed	Hierarchical scope representation	Mental model

Table 1: Predictions of how disambiguating information in (11) affects the continuation in (12)

Subject and Verb	Object	Spillover⁶	Wrap-up
The {caregivers} {caregiver} wanted	the {child} {children}	to get	some rest.

Table 2: Regions in the eye-tracking experiment

⁶ The spillover consisted of one word (if there were only two words in total after the object) or two words (otherwise).

	First pass	Go-past	Prob. of regression	Total times	Re-reading ⁷	Prob. of re-reading
Region: Subject and verb (<i>the caregiver(s) wanted</i>)						
SUBJECT:SG, OBJECT:SG	552(23)	638(34)	7(1)	737(36)	460(45)	41(4)
SUBJECT:SG, OBJECT:PL	545(25)	655(36)	8(1)	782(38)	536(58)	44(4)
SUBJECT:PL, OBJECT:SG	629(37)	773(57)	8(1)	963(57)	604(62)	55(4)
SUBJECT:PL, OBJECT:PL	592(27)	723(40)	11(1)	882(51)	680(73)	43(4)
SUBJECT: <i>different</i>	590(27)	761(45)	11(1)	956(48)	673(62)	54(4)
OBJECT: <i>different</i>	539(25)	701(52)	10(1)	732(36)	492(54)	40(1)
Region: Object (<i>the child(ren)</i>)						
SUBJECT:SG, OBJECT:SG	316(13)	382(24)	10(2)	424(21)	376(38)	29(4)
SUBJECT:SG, OBJECT:PL	336(13)	438(31)	17(3)	501(25)	387(32)	44(4)
SUBJECT:PL, OBJECT:SG	313(12)	414(27)	17(3)	460(22)	362(32)	42(4)
SUBJECT:PL, OBJECT:PL	353(14)	439(24)	12(3)	543(34)	530(67)	36(4)
SUBJECT: <i>different</i>	313(13)	513(49)	23(3)	486(27)	432(36)	42(4)
OBJECT: <i>different</i>	340(14)	511(54)	16(3)	464(22)	377(38)	34(1)
Region: Spillover (<i>to get</i>)						
SUBJECT:SG, OBJECT:SG	245(15)	522(63)	32(5)	296(18)	309(46)	21(3)
SUBJECT:SG, OBJECT:PL	304(21)	567(67)	26(4)	352(23)	273(52)	23(3)
SUBJECT:PL, OBJECT:SG	270(15)	578(106)	23(4)	341(21)	333(46)	27(3)
SUBJECT:PL, OBJECT:PL	257(15)	580(89)	34(5)	347(22)	308(38)	33(4)
SUBJECT: <i>different</i>	263(14)	758(118)	36(5)	343(21)	311(42)	29(4)
OBJECT: <i>different</i>	281(19)	525(73)	33(2)	328(20)	270(37)	21(1)
Region: Wrap-up (<i>some rest</i>)						
SUBJECT:SG, OBJECT:SG	476(30)	1275(105)	70(4)	569(39)	437(66)	21(3)
SUBJECT:SG, OBJECT:PL	459(27)	1361(96)	74(4)	599(32)	432(40)	32(4)
SUBJECT:PL, OBJECT:SG	454(26)	1538(127)	80(3)	586(41)	409(93)	32(4)
SUBJECT:PL, OBJECT:PL	475(28)	1509(154)	67(4)	602(34)	458(53)	28(4)
SUBJECT: <i>different</i>	447(28)	1544(136)	75(4)	613(38)	518(74)	32(4)
OBJECT: <i>different</i>	481(27)	1151(84)	70(2)	566(30)	367(42)	23(1)

Table 3: Means & SEs (in ms) in the second sentence

⁷ The means of total times do not equal the sum of the means of first pass and re-reading because we removed zeros from re-reading values (when total times equaled first-pass reading times).

Effect	First pass	Go-past	Prob. of regression	Total times	Re-reading	Prob. of re-reading
Region: Subject and verb (the caregiver(s) wanted)						
SUBJECT:PL	0.09($t = 1.7$)	0.09($t = 1.2$)	-0.66($z = -1.3$) ⁸	0.19 ⁸ ($t = 2.7, p = .004$) ⁸	0.21($t = 1.5$)	0.73 ($z = 3.0, p = .003$) ⁸
OBJECT:PL	-0.04($t = -0.6$)	-0.03($t = -0.5$)	-0.65($z = -1.2$) ⁸	0.002($t = 0.04$) ⁸	0.13($t = 1.1$)	0.17($z = 0.7$) ⁸
SUBJECT:PL × OBJECT PL	0.03($t = 0.4$)	0.06($t = 0.6$)	1.44($z = 1.8, p = .07$) ⁸	-0.07($t = -0.9$) ⁸	0.01($t = 0.1$)	-0.81 ($z = -2.3, p = .02$) ⁸
Region: Object (the child(ren))						
SUBJECT:PL	0.004($t = 0.07$)	0.06($t = 1.1$)	1.6 ($z = 3.0, p = .002$)	0.11 ($t = 1.9, p = .05$)	0.14($t = 1.1$)	0.74 ($z = 2.6, p = .008$) ⁸
OBJECT:PL	0.08($t = 1.7$)	0.11($t = 1.8$)	1.55 ($z = 2.8, p = .005$)	0.17 ($t = 2.2, p = .01$)	0.19($t = 1.5$)	0.74 ($z = 2.7, p = .008$) ⁸
SUBJECT:PL × OBJECT PL	0.05($t = 0.7$)	-0.04($t = -0.5$)	-2.4 ($z = -3.5, p = .0004$)	-0.08($t = -0.8$)	0.02($t = 0.1$)	-1.33 ($z = -3.2, p = .001$) ⁸
Region: Spillover (to get)						
SUBJECT:PL	0.06($t = 0.9$) ⁸	0.03($t = 0.3$)	-0.14($z = -0.3$)	0.08($t = 1.2$)	0.06($t = 0.4$)	0.67($z = 1.4$)
OBJECT:PL	0.16 ⁸ ($t = 2.2, p = .02$)	0.03($t = 0.2$)	-0.38($z = -0.1$)	0.14 ($t = 2.1, p = .04$)	-0.22($t = -1.0$)	0.43($z = 0.9$)
SUBJECT:PL × OBJECT PL	-0.19 ⁸ ($t = -2.0, p = .03$)	-0.06($t = -0.4$)	0.64($z = 1.1$)	-0.11($t = -1.3$)	0.1($t = 0.3$)	-0.17($z = -0.2$)
Region: Wrap-up (some rest)						
SUBJECT:PL	-0.03($t = -0.5$)	0.17($t = 1.9, p = .08$)	0.77 ($z = 2.6, p = .01$)	0.05($t = 0.7$)	-0.09($t = -0.5$)	0.75 ($z = 2.3, p = .02$)
OBJECT:PL	-0.07($t = -1.0$)	0.07($t = 0.8$)	0.47($z = 1.5$)	0.04($t = 0.5$)	0.13($t = 0.7$)	0.79 ($z = 2.5, p = .01$)
SUBJECT:PL × OBJECT PL	0.06($t = 0.7$)	-0.19($t = -1.4$)	-1.34 ($z = -2.9, p = .004$)	-0.04($t = -0.4$)	0.09($t = 0.4$)	-1.24 ($z = -2.7, p = .008$)

Table 4: Slopes and values of significance testing in the second sentence

⁸ The estimates are based on the model which had the maximal random-effect structure for subjects but one or more slope coefficients for items missing since any more complex model would not converge. For convergence we simplified the structure of item random effects, rather than subject random effects, because they generally accounted for the smallest variance.

	Subject	1st word after subject	2nd word after subject
CONTEXT:YES	-0.05 (t = -2.1 , p = .04)	-0.11 (t = -5 , p < .0001)	-0.08 (t = -4.4 , p = .0002)
SUBJECT:PL	-0.01(t = -0.3)	0.01(t = 1.0)	0.004(t = 0.3)
SUBJECT:PL × CONTEXT: YES	0.01(t = 0.3)	0.03(t = 1.3)	0.03(t = 1.6)
	Object	1st word after object	2nd word after object
CONTEXT:YES	-0.08 (t = -3.2 , p = .002)	-0.06 (t = -2.6 , p = .01)	0.02(t = 0.5)
SUBJECT:PL	0.000(t = -0.01)	0.001(t = 0.03)	-0.02(t = -1.1)
OBJECT:PL	0.01(t = 0.3)	0.001(t = 0.03)	-0.01(t = -0.6)
SUBJECT:PL × OBJECT:PL	0.01(t = 0.2)	0.03(t = 1.2)	0.05(t = 1.9, p = .052)
SUBJECT:PL × CONTEXT:YES	0.02(t = 0.7)	0.02(t = 0.6)	0.02(t = 0.8)
OBJECT:PL × CONTEXT:YES	0.01(t = 0.3)	0.03(t = 0.9)	0.08 (t = 2.4 , p = .02)
SUBJECT:PL × OBJECT:PL × CONTEXT:YES	-0.01(t = -0.2)	-0.04(t = -0.9)	-0.1 (t = -2.2 , p = .03)

Table 5: Slopes and values of significance testing of the effects

Supplementary file 1: Experimental items

We list all conditions only for the first item. For all the other ones, we provide only one of the early-disambiguation condition and one of the late-disambiguation conditions. The remaining conditions can be easily inferred from them.

- (1)
 - a. Every year, a different representative visited a colony. The representatives liked the colony very much.
 - b. Every year, a representative visited a different colony. The representative liked the colonies very much.
 - c. A representative visited a colony every year. The representative liked the colony very much.
 - d. A representative visited a colony every year. The representatives liked the colony very much.
 - e. A representative visited a colony every year. The representative liked the colonies very much.
 - f. A representative visited a colony every year. The representatives liked the colonies very much.
- (2)
 - a. Every morning, a different ranger spotted a lion. The rangers admired the lion for a while.
 - b. A ranger spotted a lion every morning. The ranger admired the lion for a while.
- (3)
 - a. At every summit, a different candidate discussed a social concern. The candidates wanted the concern to be taken seriously.
 - b. A candidate discussed a social concern at every summit. The candidate wanted the concern to be taken seriously.
- (4)
 - a. Every day, a different envoy met a hostage. The envoys talked to the hostage for a little while.
 - b. An envoy met a hostage every day. The envoy talked to the hostage for a little while.
- (5)
 - a. Every summer, a different travel agent visited a Greek island. The agents were on the island for a long time.
 - b. A travel agent visited a Greek island every summer. The agent was on the island for a long time.
- (6)
 - a. Every month, a different manager revised a contract. The managers rewrote the contract very speedily.
 - b. A manager revised a contract every month. The manager rewrote the contract very speedily.
- (7)
 - a. Every quarter, a different girl did a science experiment. The girls completed the experiment very quickly.

- b. A girl did a science experiment every quarter. The girl completed the experiment very quickly.
- (8) a. Every month, a different player learned a new trick. The players found the trick to be useful.
b. A player learned a new trick every month. The player found the trick to be useful.
- (9) a. Every year, a different producer endorsed an actor. The producers considered the actor to be promising.
b. A producer endorsed an actor every year. The producer considered the actor to be promising.
- (10) a. Every week, a different pupil returned a questionnaire. The pupils had answered the questionnaire only partially.
b. A pupil returned a questionnaire every week. The pupil had answered the questionnaire only partially.
- (11) a. At every event, a different author proposed a book award. The authors argued for the award very strongly.
b. An author proposed a book award at every event. The author argued for the award very strongly.
- (12) a. Every week, a different miner contacted a reporter. The miners talked to the reporter for a long time.
b. A miner contacted a reporter every week. The miner talked to the reporter for a long time.
- (13) a. Every winter, a different Neanderthal crossed a mountain. The Neanderthals knew the mountain very well.
b. A Neanderthal crossed a mountain every winter. The Neanderthal knew the mountain very well.
- (14) a. Every July, a different farmer advertized a tomato plant. The farmers had cultivated the plant very carefully.
b. A farmer advertized a tomato plant every July. The farmer had cultivated the plant very carefully.
- (15) a. Every month, a different junior lawyer disregarded a brief. The lawyers considered the brief to be a hoax.
b. A junior lawyer disregarded a brief every month. The lawyer considered the brief to be a hoax.
- (16) a. At every hearing, a different lawyer ignored an objection. The lawyers took the objection to be unfounded.
b. A lawyer ignored an objection at every hearing. The lawyer took the objection to be unfounded.

- (17) a. Every summer, a different beachfront bar attracted a famous singer. The bars paid the singer very well.
 b. A beachfront bar attracted a famous singer every summer. The bar paid the singer very well.
- (18) a. Every morning, a different broker mentioned an obvious drawback. The brokers considered the drawback to be serious.
 b. A broker mentioned an obvious drawback every morning. The broker considered the drawback to be serious.
- (19) a. Every morning, a different researcher interviewed a patient. The researchers asked the patient to return for a follow-up.
 b. A researcher interviewed a patient every morning. The researcher asked the patient to return for a follow-up.
- (20) a. Every week, a different child learned a vocabulary item. The children used the item correctly most of the time.
 b. A child learned a vocabulary item every week. The child used the item correctly most of the time.
- (21) a. In every meeting, a different author provided a crucial reference. The authors wanted the reference to be taken seriously.
 b. An author provided a crucial reference in every meeting. The author wanted the reference to be taken seriously.
- (22) a. At every summit, a different member state announced a target date. The states considered the target date to be conservative.
 b. A member state announced a target date at every summit. The state considered the target date to be conservative.¹
- (23) a. At every event, a different debater brought up a national issue. The debaters used the issue for selfish purposes.
 b. A debater brought up a national issue at every event. The debater used the issue for selfish purposes.
- (24) a. Every month, a different employee recommended a charity. The employees had worked with the charity quite a lot.
 b. An employee recommended a charity every month. The employee had worked with the charity quite a lot.
- (25) a. Every week, a different student presented a national tradition. The students considered the tradition to be very important.
 b. A student presented a national tradition every week. The student considered the tradition to be very important.

¹Due to overlook, this item had a typo (*targets date* instead of *target dates* in OBJECT: NS) in Experiment 1. The whole item was ignored in the analysis. The typo was corrected in Experiment 2.

- (26) a. In every war, a different militant attacked an American base. The militants took the base to be strategically important.
- b. A militant attacked an American base in every war. The militant took the base to be strategically important.
- (27) a. Every week, a different contestant did a physical test. The contestants passed the test with flying colors.
- b. A contestant did a physical test every week. The contestant passed the test with flying colors.
- (28) a. Every night, a different caregiver comforted a child. The caregivers wanted the child to get some rest.
- b. A caregiver comforted a child every night. The caregiver wanted the child to get some rest.
- (29) a. In every experiment, a different chimp pointed to an apple. The chimps seemed to like the apple very much.
- b. A chimp pointed to an apple in every experiment. The chimp seemed to like the apple very much.
- (30) a. At every fair, a different woman tried a hair product. The women liked the hair product very much.
- b. A woman tried a hair product at every fair. The woman liked the hair product very much.
- (31) a. Every Friday, a different visitor went to a local bar. The visitors recommended the bar very highly.
- b. A visitor went to a local bar every Friday. The visitor recommended the bar very highly.
- (32) a. At every competition, a different pianist chose a Beethoven sonata. The pianists played the sonata very well.
- b. A pianist chose a Beethoven sonata at every competition. The pianist played the sonata very well.
- (33) a. At every fair, a different store bid on a music collection. The stores had valued the collection very highly.
- b. A store bid on a music collection at every fair. The store had valued the collection very highly.
- (34) a. At every festival, a different critic praised a movie. The critics considered the movie to be innovative.
- b. A critic praised a movie at every festival. The critic considered the movie to be innovative.
- (35) a. Every week, a different newspaper discussed a pressing problem. The newspapers took the problem very seriously.

- b. A newspaper discussed a pressing problem every week. The newspaper took the problem very seriously.
- (36)
- a. Every month, a different volunteer performed a hard task. The volunteers enjoyed the task tremendously.
 - b. A volunteer performed a hard task every month. The volunteer enjoyed the task tremendously.
- (37)
- a. Every morning, a different manager met with a promising applicant. The managers wanted the applicant to return for a follow-up.
 - b. A manager met with a promising applicant every morning. The manager wanted the applicant to return for a follow-up.
- (38)
- a. Every morning, a different doctor read a chart. The doctors found the chart to be sloppy.
 - b. A doctor read a chart every morning. The doctor found the chart to be sloppy.
- (39)
- a. Every morning, a different disciple said a prayer. The disciples dedicated the prayer to the rising sun.
 - b. A disciple said a prayer every morning. The disciple dedicated the prayer to the rising sun.

Supplementary file 2: Bayesian model used to find the posterior distribution of fixed effects in Experiment 1

The BUGS / JAGS specification of a typical Bayesian model we estimated is provided below. Details of the MCMC estimation for this model: 3 chains, 40,000 iterations per chain (80,000 iterations in Experiment 2); the first 1,000 iterations were discarded (burn-in); we saved every 10th iteration only (we thinned the chains for storage convenience).

```
> model {
+ # Set up the scaling parameter for the multivariate subject ranef distribution
+ for (i in 1:4) {
+   xi_subj[i]~dunif(0, 100)
+ }
+
+ # Set up the var-covar matrix for the multivariate subject ranef distribution
+ Tau_B_raw_subj[1:4, 1:4] ~ dwish(W_subj[,], 5)
+ # W is the identity matrix, provided as data; we have 5 dofs, i.e., 4 ranefs + 1,
+ #to ensure a uniform (-1, 1) prior for the correlation between ranefs
+ Sigma_B_raw_subj[1:4, 1:4] <- inverse(Tau_B_raw_subj[,])
+ for (i in 1:4) {
+   sigma_subj[i] <- xi_subj[i]*sqrt(Sigma_B_raw_subj[i, i])
+ }
+ sigma_int_subj <- sigma_subj[1] # SD of intercepts
+ sigma_subject_pl_subj <- sigma_subj[2] # SD of slopes
+ sigma_object_pl_subj <- sigma_subj[3] # SD of slopes
+ sigma_subject_object_subj <- sigma_subj[4] # SD of slopes
+ for (i in 1:4) { for (j in 1:4) {
+ rho_subj[i, j] <- Sigma_B_raw_subj[i, j]/sqrt(Sigma_B_raw_subj[i, i]*Sigma_B_raw_subj[j, j])
+ } }
+ rho_int_subject_pl_subj <- rho_subj[1, 2]
+ rho_int_object_pl_subj <- rho_subj[1, 3]
+ rho_int_subject_object_subj <- rho_subj[1, 4]
+ rho_subject_pl_object_pl_subj <- rho_subj[2, 3]
+ rho_subject_pl_subject_object_subj <- rho_subj[2, 4]
+ rho_object_pl_subject_object_subj <- rho_subj[3, 4]
+
+ # The multivariate subject ranef distribution, i.e., modeling the correlated subject ranefs
+ for (j in 1:n_subj) {
+   B_raw_hat_subj[j, 1] <- 0
+   B_raw_hat_subj[j, 2] <- 0
+   B_raw_hat_subj[j, 3] <- 0
+   B_raw_hat_subj[j, 4] <- 0
+   B_raw_subj[j, 1:4] ~ dmnorm(B_raw_hat_subj[j, ], Tau_B_raw_subj[, ])
+   intercept_subj[j] <- xi_subj[1]*B_raw_subj[j, 1] # random intercept
+   subject_pl_subj[j] <- xi_subj[2]*B_raw_subj[j, 2]
+   object_pl_subj[j] <- xi_subj[3]*B_raw_subj[j, 3]
+   subject_object_subj[j] <- xi_subj[4]*B_raw_subj[j, 4]
+ }
+
+ # Set up the scaling parameter for the multivariate item ranef distribution
+ for (i in 1:4) {
+   xi_item[i]~dunif(0, 100)
+ }
```

```

+
+ # Set up the var-covar matrix for the multivariate item ranef distribution
+ Tau_B_raw_item[1:4, 1:4] ~ dwish(W_item[,], 5)
+ # W is the identity matrix, provided as data; we have 5 dofs, i.e., 4 ranefs + 1,
+ #to ensure a uniform (-1, 1) prior for the correlation between ranefs
+ Sigma_B_raw_item[1:4, 1:4] <- inverse(Tau_B_raw_item[,])
+ for (i in 1:4) {
+   sigma_item[i] <- xi_item[i]*sqrt(Sigma_B_raw_item[i, i])
+ }
+ sigma_int_item <- sigma_item[1] # SD of intercepts
+ sigma_subject_pl_item <- sigma_item[2] # SD of slopes
+ sigma_object_pl_item <- sigma_item[3] # SD of slopes
+ sigma_subject_object_item <- sigma_item[4] # SD of slopes
+ for (i in 1:4) { for (j in 1:4) {
+   rho_item[i, j] <- Sigma_B_raw_item[i, j]/sqrt(Sigma_B_raw_item[i, i]*Sigma_B_raw_item[j, j])
+ } }
+ rho_int_subject_pl_item <- rho_item[1, 2]
+ rho_int_object_pl_item <- rho_item[1, 3]
+ rho_int_subject_object_item <- rho_item[1, 4]
+ rho_subject_pl_object_pl_item <- rho_item[2, 3]
+ rho_subject_pl_subject_object_item <- rho_item[2, 4]
+ rho_object_pl_subject_object_item <- rho_item[3, 4]
+
+ # The multivariate item ranef distribution, i.e., modeling the correlated item ranefs
+ for (j in 1:n_item) {
+   B_raw_hat_item[j, 1] <- 0
+   B_raw_hat_item[j, 2] <- 0
+   B_raw_hat_item[j, 3] <- 0
+   B_raw_hat_item[j, 4] <- 0
+   B_raw_item[j, 1:4] ~ dnorm(B_raw_hat_item[j, ], Tau_B_raw_item[, ])
+   intercept_item[j] <- xi_item[1]*B_raw_item[j, 1] # random intercept
+   subject_pl_item[j] <- xi_item[2]*B_raw_item[j, 2]
+   object_pl_item[j] <- xi_item[3]*B_raw_item[j, 3]
+   subject_object_item[j] <- xi_item[4]*B_raw_item[j, 4]
+ }
+
+ # Model the resid. sd independently
+ sigma_res~dunif(0, 100) # Residual standard deviation
+ tau_res <- 1/(sigma_res*sigma_res)
+
+ #Priors for fixed effects
+ alpha ~ dnorm(0, 0.001) # intercept
+ subject_pl ~ dnorm(0,0.001)
+ object_pl ~ dnorm(0,0.001)
+ subject_object ~ dnorm(0, 0.001)
+
+ #Likelihood
+ for (i in 1:n_obs) {
+   mu_obs[i] <- #Fixefs
+   alpha +
+   subject_pl * SUnumber[i] +
+   object_pl * DOnumber[i] +
+   subject_object * SUnumber[i] * DOnumber[i] +
+   #Subject Ranefs
+   intercept_subj[subjects[i]] +
+   subject_pl_subj[subjects[i]] * SUnumber[i] +
+   object_pl_subj[subjects[i]] * DOnumber[i] +

```

```
+          subject_object_subj[subjects[i]] * SUnumber[i] * DOnumber[i] +
+          #Item ranefs
+          intercept_item[items[i]] +
+          subject_pl_item[items[i]] * SUnumber[i] +
+          object_pl_item[items[i]] * DOnumber[i] +
+          subject_object_item[items[i]] * SUnumber[i] * DOnumber[i]
+ y[i]~dnorm(mu_obs[i], tau_res)
+ }
+ }
```