What do German quantifiers tell us about semantic theories?

Abstract

In this paper we use German data to evaluate configurational and multi-factor approaches to quantifier scope. Configurational theories derive scope relations syntactically at the level of Logical Form; semantic and pragmatic factors are either built into the syntactic representation or ignored. By contrast, multi-factor approaches consider syntactic, semantic and pragmatic properties of quantifiers as multiple constraints affecting quantifier scope. In particular, we tested predictions of the configurational theory on quantifier scope in German by Frey (1993) and of the multi-factor theory by Pafel (2005). We tested these fundamentally different approaches in a series of picture verification experiments studying scope preferences in doubly quantified German sentences. The results show that at least three factors affect scope preferences. Our findings are neither fully consistent with Frey’s configurational approach nor with Pafel’s multi-factor theory. Neither theory made correct predictions for German doubly quantified sentences with a subject-before-object word order. For object-before-subject sentences, however, the experimental data by and large support the predictions of Pafel (2005)’s multi-factor approach.

1 Introduction

Intuitive judgments are the main data source semanticists rely on to formulate their theories. Often, however, semantic judgments are not very clear and preferences are gradient rather than clear-cut and categorical. In particular, scope ambiguity of quantifiers is a phenomenon where intuitions tend to be shaky.

The aim of the present paper is to systematically investigate the influence of some factors that have been claimed to influence quantifier scope preferences (e.g. Ioup, 1975; Kuno, 1991; Beghelli & Stowell, 1997; Pafel, 2005):

1. the linear order of quantifiers
2. distributivity
3. discourse anaphoricity

These three factors have played an important role in various theories of quantifier interaction. In particular linear order (e.g. Reinhart (1983), Frey (1993), Kuno (1991), Pafel (2005)) and distributivity of quantifiers (e.g. Beghelli & Stowell (1997), Szabolcsi (1997), Kuno (1991), Pafel (2005)) have commonly
been claimed to affect scope. The same holds for the discourse status of quantifying expressions. Quantifiers involving a partitive construction such as *each of these* make explicit reference to a contextually given restrictor set. D-linking, which is related to discourse anaphoricity in that it also presupposes a contextually given restrictor set, has been discussed in the generative literature on wh-operators (e.g. Pesetsky (1987), Cinque (1990)), where it has been argued that d-linked phrases can be interpreted in situ whereas non-d-linked phrases have to undergo movement at LF. In a similar vein, Kuno (1991) and Pafel (2005) both predict that quantifiers with a discourse anaphoric restriction tend to take wide scope more easily than quantifiers without.

The factors have been captured in different ways in different types of theoretical approaches. In this paper we contrast syntactic approaches to quantifier scope with multi-factor theories. The latter take into account syntactic as well as non-syntactic factors without drawing any qualitative distinctions between them. We will completely ignore semantic and pragmatic approaches to the resolution of quantifier scope ambiguity (see among others Hendricks (1993), Barker (2002)). Even with this limitation in mind we cannot do justice to all existing scope theories but will focus on a very limited set of proposals dealing with quantifier scope in German (for general reviews see Kiss (2006), Ruys & Winter (2010), Szabolcsi (2010) and Steedman (2011)).

1.1 Configurational Accounts

Configurational accounts treat scope ambiguities as a kind of syntactic ambiguity. Montague (1973) was the first to propose a syntactic rule *quantifying in* which was specifically designed to generate inverse scope readings of multiply quantified sentences. Given different syntactically disambiguated structures, semantic interpretation can proceed straightforwardly in a compositional fashion. Similarly, rather than taking the surface structure as input for semantics, May (1977) argues that syntax continues the derivation until Logical Form (LF), the level at which interpretation takes place. Quantifiers move covertly to a position from which they c-command their scope. This mechanism is called *Quantifier Raising (QR)* and has been adopted in several modified versions (May 1985, Hornstein & Weinberg 1990, Reinhart 1995, Heim & Kratzer 1998, Fox 2003). In unconstrained versions of QR the prediction is that a sentence containing *n* quantifiers should have *n*! readings. Minimalist versions of configurational theories differ in terms of the underlying grammatical architecture but share the assumption that semantic interpretation can target unpronounced constituents of syntactic representations with multiple occurrences of quantifiers in feature-driven derivations (?, Fox 2003).

Besides QR, another covert syntactic transformation has played an important role in configurational accounts: the reconstruction of quantifiers into their base positions. A proposal along these lines for German is Höhle (1991). For German, it is assumed that scope reconstruction rather than QR is the mechanism underlying inverse scope phenomena (see Sauerland (2003)). Using this mechanism scope readings can be directly read off from surface structure. In this paper, we will focus on the scope theory proposed in Frey (1993) devel-
oped to account for operator scope in German. Consider the following German examples with verum focus (Höhle 2002), indicated by capitalization of the verb.

\[(1)\] a. Fast jeder Lehrer LOBTE mindestens einen Schüler.
   Almost every teacher praised at least one student.

   b. Mindestens einen Schüler LOBTE fast jeder Lehrer.
   At least one student\textsubscript{object} praised almost every teacher\textsubscript{subject}.

   ‘Almost every teacher praised at least one student.’

\[(2)\] a. [Fast jeder Lehrer\textsubscript{1} [lobte\textsubscript{2} [t\textsubscript{1} mindestens einen Schüler t\textsubscript{2}]])

   b. [Mindestens einen Schüler\textsubscript{1} [lobte\textsubscript{2} [fast jeder Lehrer t\textsubscript{1} t\textsubscript{2}]])

   The corresponding syntactic representations involve two movement operations. The verb is moved to the V2 position and one of the quantificational DPs is moved to the prefield: the subject in (1-a), and the object in (1-b). There is an intuitive difference between (1-a) and (1-b) regarding: While (1-b) is ambiguous, (1-a) seems to only allow for a linear scope construal. Frey proposes a scope principle to account for this difference; here is a simplified version, which will do for our purposes (cf. Frey (1993)).

\[(3)\] Scope Principle (simplified, informal version): A quantifier $\alpha$ can take scope over another quantifier $\beta$ if $\alpha$ and $\beta$ are in the same syntactic domain and $\alpha$ syntactically commands $\beta$ or its trace.

In the syntactic representation of (1-a) the subject DP \textit{almost every teacher} c-commands the object quantifier \textit{at least one student}, but not vice versa. This is different in (1-b) where the fronted object quantifier takes command over the subject quantifier, but the subject quantifier c-commands the trace $t_1$ of the object quantifier. This way, the configuration in which the two quantifiers appear posits constrains on the available scope readings.

The theory comes with two important restrictions. Firstly, it only accounts for the interpretation of truly quantificational phrases, i.e. DPs that are not ambiguous between quantificational and other interpretations such as indefinites (Fodor & Sag 1982), which are known to allow for specific, referential interpretations with exceptional wide scope. This is why the above examples used German equivalents of \textit{at least one} instead of indefinites such as \textit{a} or \textit{some}. The same goes for group denoting DPs such as \textit{all} or even \textit{every}, which – in one way or the other – make reference to plural objects. For this reason the above examples used German equivalents of \textit{almost every} instead of plain \textit{every}. Secondly, Frey observes that prosody may make interpretations available that are impossible otherwise (see also Krifka (1998)). To avoid stress on any of the quantificational DPs he controls for intonation by always choosing verum focus. If these two restrictions are met, his configurational account predicts scope ambiguity to be absent in German sentences such as (1-a) with subject-before-object word order.

Let us now turn to sentences with object-before-subject word order. As just shown, this syntactic configuration opens up the possibility of inverse scope.
This does not mean, however, that all sentences of this type should be fully ambiguous. Other, non-syntactic factors may still influence scope preferences. We can think of them as a filter on the set of possible readings that are grammatically licensed in a given configuration. In principle, it is thus even possible that one of the syntactically licensed readings is completely ruled out by some other factor or a combination of factors.

Indeed, it has been observed that quantifying expressions are not uniform with respect to their tendency to take wide scope (Szabolcsi 2010). Some configurational accounts have directly built influences of prima facie non-syntactic factors into their representations. Beghelli & Stowell (1997) have developed a minimalist version of QR in terms of feature checking. Quantifiers take scope from different positions in a split-CP structure. In their theory distributive phrases like each boy have to check their features at LF at a projection DistP high up in the tree while non-distributive phrases like all boys may remain low. This accounts for the influence of distributivity on a quantifier’s tendency to take wide scope. What at first sight seem to be lexical semantic distinctions between different determiners are thus claimed to be configurational differences between logical forms. In principle this strategy can be applied to many factors that affect the scopal behavior of quantifying expressions. Since no worked-out proposal of this kind exists for German, we will limit our discussion to the configurational account of (Frey 1993).

An important consequence of incorporating different scope factors directly into the hierarchical representations is that scope factors should interact in an asymmetrical fashion. If a quantifier feature has to be checked in a high position at LF it should not matter whether the other quantifier involves feature checking at some lower hierarchical level. We will come back to this point when discussing the purely additive effects observed for manipulations of the different scope factors in Experiments 2 and 3 reported below.

1.2 Multi-factor Theories

Multi-factor theories assume a rather indirect relationship between syntax and semantics. Syntactic constellation is just one of a number of factors which determine the relative scope of quantifiers. Unlike configurational accounts, multi-factor theories aim at modeling the exact distribution of readings observed for multiply quantified sentences including scope preferences. First and foremost, they therefore strive for descriptive adequacy. These theories comprise very early accounts like those of Kroch (1974), Ioup (1975) and VanLehn (1978) but also the somewhat more recent proposals by Kuno (1991) and Pafel (2005). According to the latter two accounts a genuinely syntactic factor (linear order) plus a series of other factors such as distributivity and discourse-anaphoricity of quantifiers influence relative scope. In Pafel’s theory the relative scope of German multiply quantified sentences is predicted employing a linear additive model which uses weighted factors to determine the scope value of a given quantifier and rank the quantifiers with respect to their values. Multi-factor theories claim that the scope determining factors are independent from each other and that they add up in a purely cumulative fashion.
To illustrate how scope preferences are derived in a multi-factor theory, we come back to the examples (1-a) and (1-b) from above. The following predictions are derived from Pafel (2005)’s theory on relative scope in German with one slight modification. Pafel uses threshold values, i.e. difference scores between quantifiers’ ‘scope potential’, to distinguish between ambiguous and unambiguous sentences. We do not employ thresholds here because existing experimental work suggests that their use leads to a loss of explained variance relative to a model without thresholds (Bott & Radó 2007). The only relevant factors with respect to (1-a) and (1-b) from Pafel’s model are linear order, grammatical function, and distributivity. Other factors such as discourse anaphoricity (‘discourse binding’), thematic properties of arguments related to psych verbs, intonation and so forth do not play a role here. Each quantifier is evaluated with respect to these dimensions by computing its respective scope value (s-val) according to a linear additive model using factor weights. If a quantifier in the prefield precedes another one (1st, for short) a value of 1.5 is added, if it functions as subject (subj., for short) a value of 1 is added, and if it is distributive (dist., for short) its scope value is increased by 1. Thus, the following scope values are derived.

(4) a. \[ s\text{-val (almost every teacher)} = 1st \& subj. \& dist. = 1.5 + 1 + 1 = 3.5 \] (ex. (1-a))
   \[ s\text{-val (at least one student)} = 0 \] (ex. (1-a))

b. \[ s\text{-val (at least one student)} = 1st = 1.5 \] (ex. (1-b))
   \[ s\text{-val (almost every teacher)} = subj. \& dist. = 1 + 1 = 2 \] (ex. (1-b))

Comparing the respective scope values shows that the predicted readings are more or less the same as those predicted by Frey (1993). For (1-a), the scope value of almost every teacher is much higher as the one for at least one student which corresponds to a very strong preference for linear scope, whereas the much closer values for (1-b) indicate a more or less ambiguous sentence with possibly a slight preference for inverse scope. However, if we change quantifiers the two theories make rather different predictions from each other. Consider (5).

(5) a. Genau ein Lehrer LOBTE fast jeden dieser Schüler.
   Exactly one teacher praised almost every of these students.

b. \[ s\text{-val (exactly one teacher)} = 1st \& subj. = 1.5 + 1 = 2.5 \]
   \[ s\text{-val (almost every of these students)} = dist. \& d\text{-ana.} = 1 + 1 = 2 \]

Thus in Pafel’s multi-factor theory a small change of quantifiers should result in a rather large change in scope distributions. (5) with subject-before-object word order is predicted to be more or less scope ambiguous even though a linear scope construal should be preferred. For doubly quantified sentences with an object-before-subject word order the prediction is again that factors other than the syntactic configuration should have a strong influence on scope preferences.

Consider (6).

---

1And even more so in Pafel’s original model with thresholds where (1-a) is predicted to unambiguously exhibit surface scope.

2Discourse anaphoricity (d-anaphoric) has a factor weight of 1.
While (6-a) is predicted to show a very strong preference for linear scope, in (6-c) inverse scope should actually be preferred over the linear reading. The discussion of these examples illustrates that the aim of multi-factor accounts is rather different from configurational theories: They try to account for subtle differences in scope preferences at a rather descriptive level. Unlike configurational accounts multi-factor theories are far less concerned with identifying a single mechanism that explains why particular readings are available or unavailable. Most notably, their predictions lend themselves to psycholinguistic studies because they can be translated into quantitative predictions for experiments. We conducted a study addressing the following questions.

1. Do doubly quantified sentences of the type exemplified in (5) with subject-before-object word order have unambiguous linear scope as predicted by Frey (1993), or are they ambiguous as predicted by Pafel (2005)?

2. Do we find evidence for the scope distributions predicted by Pafel’s multi-factor theory for object-before-subject sentences of the types illustrated in (6)?

3. If the scope factors claimed relevant for quantifier scope do in fact show the predicted influences, how do they interact? Do we find evidence for purely additive effects as essentially predicted by additive linear models of the sort just outlined? Or, do we rather find over-additive effects as may be expected under a configurational account that encodes non-syntactic factors at dedicated positions in the LF (Beghelli & Stowell 1997)?

The first two questions will be addressed in our first experiment, the third question will be taken up in the second and the third experiment. Before going into the experimental study in section 3, we will first relate our research questions to existing psycholinguistic literature on quantifier scope ambiguity, then in section 2 we will discuss methodological considerations important for assessing quantifier scope preferences.
1.3 Psycholinguistic Studies on Quantifier Scope

Multi-factor models have been quite prominent in the psycholinguistic literature on the processing of quantifier scope. We cannot go into detail here but refer the reader to the overview in Tunstall (1998, chap. 2) and the more recent studies by Paterson, et al. (2008) and Brasoveanu & Dotlacil (2015). It has been repeatedly claimed that the language processor simultaneously relies on different sources of information to generate the possible readings of multiply quantified sentences (e.g. Kurtzman & MacDonald (1993), Filik, et al. (2004), Paterson et al. (2008)). Unfortunately, the existing psycholinguistic studies cannot easily be related to the scope theories reviewed above.

One reason is that many of the existing studies used suboptimal disambiguations. In Kurtzman & MacDonald (1993), Tunstall (1998) and Dwivedi (2013) as well as in the Filik et al. (2004) and Paterson et al. (2008) studies a potentially scope ambiguous sentence with a universal and an existential quantifier like every kid climbed some tree was combined either with a singular or with a plural continuation (e.g. this tree/these trees was/were full of apples) which was intended to disambiguate the scope. However, it has been often noted both in the semantic (e.g. Kempson & Cormack (1981)) and in the processing literature (e.g. Tunstall (1998)) that the $\exists\forall$-reading entails the $\forall\exists$-reading. Therefore, the singular continuation is compatible with both interpretations and does not achieve real disambiguation. This can be illustrated by combining an unambiguously $\forall\exists$-sentence with a singular continuation in (7).

(7) Every child is such that it climbed some tree. This tree was full of apples.

More generally, the method used to assess whether a particular reading is available must not introduce a distorting bias. In the experiments presented here, we will therefore use a picture verification task that has been shown to yield reliable and valid results (reference removed).

Second, to test the theoretical approaches outlined above we need to systematically investigate the interplay of the various factors that have received attention in the semantic literature on quantifier scope. Up to now, none of the existing psycholinguistic studies has manipulated linear order, distributivity and discourse anaphoricity in tandem using the same procedure and items. As outlined above, to evaluate multi-factor theories it is crucial to test their prediction of purely additive scope influences when systematically manipulating different scope factors at the same time.

However, there is a body of experimental work examining these factors separately. The potential influence of linear order has been investigated but the findings are mixed. While VanLehn (1978), Fodor (1982), Gilden (1991), Kurtzman & MacDonald (1993), Tunstall (1998), Anderson (2004) and Paterson et al. (2008) reported effects of linear order and/or c-command relations, Ioup (1975), Catlin & Micham (1975), and Micham, et al. (1980) found no effects of surface order but only of grammatical function. Distributivity has not received much attention except for Tunstall (1998), Bott & Radó (2009), and Radó & Bott
(2012), who provided initial evidence for its influence as a scope factor\textsuperscript{3}. In a recent study, Brasovaeau & Dotlacil (2015) found that each led to significantly higher proportions of inverse scope interpretations than every. To our knowledge, the influence of discourse anaphoricity with respect to quantifier scope has not been tested experimentally in adults. However, experimental work on language acquisition by Musolino & Gualmini (2004) suggests that discourse anaphoric partitive quantifiers can be more easily interpreted with inverse scope than non-partitive quantifier phrases.

Finally, almost all of these studies investigated quantifier scope in English. Because English has strict SVO word order, in an active declarative sentence the subject always precedes the object. Thus the effect of linear order/c-command relations and of the grammatical function of the quantifiers cannot be easily teased apart. In other languages word order is much less constrained and these factors can be better distinguished. We therefore chose German, where the direct object can precede the subject\textsuperscript{4}. This is illustrated in (8-a) and (8-b).

(8)  
\begin{enumerate}[a.]
\item Jeder Lehrer lobte genau einen Schüler.
Every teacher\textsubscript{subject} praised exactly one student\textsubscript{dir. object}.
\item Genau einen Schüler lobte jeder Lehrer.
Exactly one student\textsubscript{dir. object} praised every teacher\textsubscript{subject}.
\end{enumerate}

However, there is one line of study that is closely related to the questions addressed in the present paper. Bott & Schlotterbeck (2012) investigated whether German object-before-subject and subject-before-object sentences both allow inverse scope readings. They were particularly interested in the readings that become available during the online interpretation of doubly quantified sentences. Their study provided evidence that subject-before-object sentences differ from object-before-subject sentences in that only the latter gave rise to online effects of scope ambiguity. The final interpretations as measured in an offline task, however, indicated ambiguity in both types of constructions. Furthermore, Bott & Schlotterbeck (2012) only manipulated the configuration that the quantifiers appeared in, they did not investigate the interactions with other, non-syntactic scope factors.

To sum up, we think that the existing studies are not sufficient to evaluate the scope accounts discussed above. Nonetheless, they provide initial support that scope is influenced by different factors. In the following we will present three picture verification experiments that were designed to test the two scope theories we focus on: Frey (1993)’s configurational account and the multi-factor approach by Pafel (2005).

\textsuperscript{3}But see Filik et al. (2004) and Paterson et al. (2008), who did not find stronger effects for each than for every. In their studies, however, distributivity was only manipulated across experiments.

\textsuperscript{4}Although object-subject word order has been shown to be more complex in processing than subject-object order (see e.g. Hemforth & Konieczny (1999) and the papers therein). Note that we are not making any claims about processing but are only interested in the final interpretation.
2 Methodological Aside

To determine scope preferences, we have to relate a potentially scope ambiguous construction to each of its potential meanings and measure how well they fit. To this end, we used a picture verification task and had participants judge how well doubly quantified sentences match their disambiguations. All sentences contained a universal quantifier (jed- (every) or alle (all); ∀) and genau ein (exactly one) (‘∃!’). Choosing a non-monotone quantifier like exactly one was necessary in order to provide proper disambiguations for the sentences (see, e.g., the methodological discussion in Ruys & Winter (2010)). The disambiguations were set diagrams only consistent with one scope reading, one model disambiguating towards $∃∀$, the other disambiguating towards $∀∃$. Figures 1 and 2 show two disambiguating models for sentence (9). Set diagrams of the sort used here have been successfully employed in a number of experimental studies on scope (see Bott & Radó (2007) for a cross-methodological study investigating the reliability and validity of the method, and Gillen (1991), Jackson & Lewis (2005), Bott & Schlotterbeck (2012), Robaldo, et al. (2014), and Bott & Schlotterbeck (2015) for applications).

(9) Exactly one student was praised by every teacher.

The model on the right is true on the $∃∀$ reading since there is exactly one student who was praised by all the teachers. It is false on the $∀∃$ reading because one student was praised by more than one teacher. By contrast, the model on the left is true on the $∀∃$ reading because for each teacher there is exactly one student whom he praised. It is false on the $∃∀$ reading because there is no student who was praised by all the teachers. The $∃∀$ diagrams were always of the kind in Figure 2, that is, all of them contained at least one additional line, which made the $∀∃$ reading false.

We are now able to operationalize what should count as a scope preference (criterion to determine scope preference): reading A is preferred over reading B if the A-disambiguation is rated better than the B-disambiguation; if the ratings are equal, the sentence is ambiguous.

As mentioned in the introduction, Frey (1993)’s theory is limited in important respects. It only applies to a small number of quantificational expressions because other expressions are ambiguous between specific/referential and quantificational expressions (i.e. ‘indefinites’) or between collective and quantificational expressions (i.e. ‘plurals’). The quantificational expressions tested in our study are exactly one (of these), every (one of these) and all (of these). The comparison between jed- (every) and alle (all) is important to test for distribu-
tivity effects. However, the results for conditions with alle will not be used to evaluate Frey’s theory but will only become relevant when testing the additivity of scope factors. Crucially, the quantifiers genau ein (exactly one) and jed-(every) are unambiguously quantificational, i.e. they do not allow referential and/or collective interpretations. They can thus be used to test Frey’s configurational account. Let’s first consider the possibility of referential interpretations of exactly one.

(10) a. Peter glaubt, dass ihn (genau) eine Vegetarierin liebt.
Peter believes that he is loved by exactly one/a vegetarian.

b. Falls (genau) einer meiner Onkels dieses Jahr stirbt, erbe ich ein Vermögen.
If (exactly one)/some of my uncles dies this year, I’ll inherit a fortune.

(10-a) employs a test used by Frey (1993). The sentence without exactly, i.e. indefinite eine (a) allows for a de re reading where Peter believes that he is loved by some woman without knowing that she is a vegetarian. This reading is not possible with exactly one, mirroring Frey’s observation for the monotone quantifier midestens ein- (at least one). In other words, genau ein cannot take exceptional wide scope over believe. (10-b) (cf. Fodor & Sag (1982)) shows that exactly one cannot outscope the if-clause. The sentence allows only for the pragmatically marked interpretation that there is a strange inheritance rule according to which I inherit a fortune only if one but not more than one of my uncles dies (no matter which one). The specific interpretation, easily available without genau, becomes impossible once we add exactly.

The same goes for the unavailability of collective interpretations of jed- (every). Jed- cannot co-occur with collective predicates such as surround, form a line, or run apart resulting in odd interpretations in (11-a)–(11-c). As mentioned above Frey additionally dismisses unmodified jed- and only uses fast jed- (almost every). The reason is that in his view unmodified jed- makes reference to a given restrictor set, as evidenced by strong readings of unmodified examples (Dalrymple, et al. 1998, Schlotterbeck & Bott 2013). Pafel (2005), however, observes that basically the same readings are available for jed- and fast jed- in (11-d). The most natural interpretation for the sentence is that for (almost) each pair of individuals from the domain the two individuals differ in their smell. This interpretation is available with or without almost, contrary to Frey’s proposal. In order to avoid problems due to the inherent vagueness of almost we therefore used unmodified jed- in the experiments reported below.
Finally, in order to address the first research question stated above, it is crucial to define a criterion that allows us to decide whether doubly quantified sentences with consistently low ratings for the inverse interpretation are still in principle ambiguous or only have the linear reading. We will therefore compare potentially scope ambiguous sentences such as (5), (6-a), and (6-c) with scope disambiguated baseline control conditions in which quantifiers are separated by a clause boundary (see also Bott & Schlotterbeck (2012) for the same rationale). This is illustrated in the following English translations of the three experimental conditions to be tested in Experiment 1.

(12) a. Exactly one teacher is such that he praised each of these students.
b. Each of these students is such that he was praised by exactly one teacher.
c. Exactly one student is such that he was praised by each of these teachers.

The comparison with unambiguous constructions allows us to operationalize what we mean by saying that a sentence allows for, or rather doesn’t allow for some scope reading (criterion to determine scope ambiguity): Sentence S allows for scope reading A, iff the picture disambiguating for A is judged more acceptable when paired with S than with the appropriate disambiguated baseline control S’. If there is no difference between S and S’ with respect to their distributions of acceptability scores for the potential readings, S is unambiguous.

2.1 Pretests

Before testing potentially scope ambiguous sentences we conducted two pretests that confirmed that neither did the disambiguating pictures induce a typicality bias nor did the sentence materials introduce a plausibility bias.

Pretest 1: We pretested the disambiguating pictures with scope disambiguated sentences to make sure that they do not introduce any bias towards one of the readings. The situations depicted by the set diagrams may be more or less typical for the reading they are instantiations of. For instance, one may consider the one-to-one mapping between teachers and students in Figure 1 a rather typical situation for the $\forall \exists !$ reading, whereas the set diagram with the extra line in Figure 2 might instantiate a possibly less typical situation for the $\exists \forall$ reading. As a consequence, participants may therefore reject $\exists \forall$ disambiguations such as the one in Figure 2 even in cases when they have the $\exists \forall$ reading in their mind.
This would result in a bias towards accepting ∀∃-diagrams more often than ∃∀-diagrams (see Bott & Schlotterbeck (2012) and Bott & Schlotterbeck (under review) for experimental evidence that this can happen in a timed version of the task, but that – contrary to what we suggest above – typicality effects mainly affected ∀∃ disambiguations (of a slightly different type) and not the ∃∀ disambiguations).

We pretested the disambiguating set diagrams with scope disambiguated sentences. The 24 items to be tested in Experiments 2-3 were disambiguated by putting a clause boundary between the quantifiers. (13-a) is an unambiguously ∃∀ sentence, whereas (13-b) is unambiguously ∀∃. The complete list of items is provided in Appendix B.

(13) a. Für genau einen Aufsatz gilt, dass ihn alle Studenten gelesen haben. "Exactly one paper is such that every student read it."
b. Für jeden Studenten gilt, dass er genau einen Aufsatz gelesen hat. "Every student is such that he read exactly one paper."

The sentences were combined with either a ∀∃- or a ∃∀-diagram (see Figure 1 and Figure 2) according to a 2 x 2 factorial design with sentence (∀∃- vs. ∃∀-sentence) and diagram (∀∃- vs. ∃∀-diagram) as within factors. Using a latin square design, the items and 36 distractors were distributed over four lists. 24 participants judged whether the sentence matched the picture or not. Table 1 presents percent “yes” answers. The results show that the sentence-picture pairs were only judged as matching when they had the same scope. ANOVAs revealed a significant interaction between sentence and diagram (F1(1,23) = 267.3; p < .01; F2(1,23) = 1103.5; p < .01) while the two main effects of sentence and of diagram were not significant. Thus, the set diagrams used for disambiguation provide proper disambiguation and do not artificially induce a bias towards either of the readings (for further methodological considerations concerning the reliability and validity of the experimental task used in our experiments see Bott & Radó, 2007).

<table>
<thead>
<tr>
<th></th>
<th>∀∃-diagram</th>
<th>∃∀-diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>∀∃-sentence</td>
<td>95.1%</td>
<td>6.3%</td>
</tr>
<tr>
<td>∃∀-sentence</td>
<td>14.6%</td>
<td>88.9%</td>
</tr>
</tbody>
</table>

Table 1: percent “yes”-answers in the pretest.

Pretest 2: We conducted a second pretest to make sure that the sentence materials did not introduce a pragmatic bias towards either scope reading. Ruys & Winter (2010), for instance, explicitly discuss the role of biases due to plausibility in examples like an apple in every basket where the linear interpretation is highly implausible. To guarantee that the sentence materials were equally plausible in the ∃∀ and the ∀∃ reading we obtained plausibility norms for the experimental sentences. For each item two questions (∃∀ vs. ∀∃) were constructed, for instance German variants of how plausible is it that all students read one and the same novel? vs. how plausible is it that each student read exactly
one novel but different students read different novels? We used a latin square to construct two counterbalanced lists with twenty-four questions each. The items from Experiments 2-3 were rated by twenty-four German native speakers on a seven-point scale (1 = very plausible, 7 = very implausible). \( \exists ! \forall \) readings had a mean plausibility rating of 2.92 and \( \forall \exists ! \) was rated 2.80. The difference between the readings was not significant (\( t_{1,2} < 1 \)). Thus there was no bias towards one particular scope reading, both interpretations were equally plausible.

3 Frey (1993) vs. Pafel (2005) (Exp. 1)

When introducing the accounts by Frey (1993) and by Pafel (2005) in the introduction we picked out three doubly quantified constructions for which the theories make fairly different predictions. These were the three constructions tested in the first experiment.

3.1 Methods

3.1.1 Materials

We constructed 36 doubly quantified sentences in three sentence conditions each. Here is a sample item.

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

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

c. Genau ein Lehrer lobte jeden dieser Schüler...
    'Exactly one teacher praised each of these pupils...'

Conditions (14-a) and (14-b) are object-before-subject sentences. They differ, however, with respect to the relative factor weights of the quantifiers. While in example (14-a) based on Pafel (2005)’s theory we expect full ambiguity with a slight tendency of the distributive subject quantifier \textit{each teacher} to take scope over the topicalized object \textit{exactly one student}, in (14-b) the cumulative factor weights of linear order, distributivity and discourse-anaphoricity of \textit{exactly one teacher} work against the subject quantifier \textit{each teacher} and we predict the sentence to be strongly biased towards linear scope. According to Pafel (2005)’s theory changing the order of the quantifiers in (14-c) should yield an ambiguous sentence, this time slightly biased towards the linear interpretation. This is because the sentence-initial quantifier is the subject whereas the second quantifier is distributive and discourse-bound. The magnitude of the preference should be exactly the same as in the first object-before-subject condition (14-a), but go in the opposite direction. The factor weights and predicted scope values based on Pafel (2005) for each of the three constructions in (14) are summarized in (15).
(15) a. Genau einen Schüler lobte jeder Lehrer voller Wohlwollen. 
\[ \exists! \: 1.5 \ (1st) = 1.5 \]
\[ \forall \: 1 \ (subj.) + 1 \ (dist.) = 2 \]
\[ \triangleright \text{slight preference for inverse scope} \]

\[ \exists! \: 1 \ (subject) = 1 \]
\[ \forall \: 1.5 \ (1st) + 1 \ (dist.) + 1 \ (d-ana.) = 3.5 \]
\[ \triangleright \text{strong preference for linear scope} \]

c. Genau ein Lehrer lobte jeden dieser Schüler voller Wohlwollen. 
\[ \exists! \: 1.5 \ (1st) + 1 \ (subj.) = 2.5 \]
\[ \forall \: 1 \ (dist.) + 1 \ (d-ana.) = 2 \]
\[ \triangleright \text{slight preference for linear scope} \]

However, if scope ambiguity is only possible in the object-before-subject sentences the subject-before-object sentence (14-c) should have unambiguously linear scope. To test this prediction by Frey (1993) we added three scope unambiguous controls with quantifiers appearing in clause bounded position. If the potentially ambiguous subject-before-object construction (14-c) should turn out to be indistinguishable from a corresponding scope unambiguous control condition we can legitimately take it to be unambiguous. Consider the sample item in the three control conditions in (16).

(16) a. Für genau einen Schüler gilt: ihn lobte jeder Lehrer... 
For exactly one pupil holds: him praised each teacher...
'Exactly one pupil is such that he was praised by each teacher...'

b. Für jeden dieser Schüler gilt: ihn lobte genau ein Lehrer... 
For each these pupils holds: him praised exactly one teacher...
'Each of these pupils is such that he was praised by exactly one teacher...'

c. Für genau einen Lehrer gilt: er lobte jeden dieser Schüler... 
For exactly one teacher holds: he praised each these pupils...
'Exactly one teacher is such that he praised each of these pupils...'

Each of the six sentence conditions was paired with a picture that was only compatible with the \[ \forall \exists! \] or the \[ \exists! \forall \] reading. This yielded a total of twelve sentence-picture pair conditions according to a 3 \( (\text{construction: os1 vs. os2 vs. so}) \times 2 \ (\text{ambiguity: amb. vs. disamb.}) \times 2 \ (\text{disambiguation: linear vs. inverse}) \) factorial design. The disambiguating pictures in the present experiment were of exactly the same kind as in Experiments 1 and 2. Sample pictures for all three construction types are given in Figures 3–5. The \[ \exists! \forall \]-picture was always the same in the os1, the os2 and the so sentences. The \[ \forall \exists! \] pictures in the os1 order were just the mirror image of those used in the os2 and so orders. We added 61 filler sentence-picture pairs (15 false) and constructed 12 lists using a latin square design.
3.1.2 Procedure and Participants

The experiment was programmed using WebExp 2 (Mayo, et al. 2006). After reading written instructions, participants first completed a short practice session consisting of ten trials. That was followed by the experiment in one block. Sentence-picture pairs were presented in an individually randomized order. Judgments were provided on a seven point scale. We measured both judgments and judgment times.

48 students at Tübingen University (mean age 24.7 years, range 20–33 years; 32 female) participated in the study for a payment of 5€. Participants were tested individually in a quiet computer pool at the Department of Modern Languages. An experimental session took about 30 minutes.

3.2 Results and Discussion

The ratings were normalized by computing z-scores for each participant. Mean judgments are shown in Figure 6. We will report separate 2 (ambiguity)
x 2 (disambiguation) repeated measures ANOVAs for the three constructions. The so conditions provide evidence for the possibility of inverse scope in this construction type. Even though the linear interpretation was strongly preferred over the inverse reading (mean z-scores: 0.41 vs. -0.89) this preference was significantly weaker than the corresponding difference in the unambiguous conditions (mean z-scores: 0.63 vs. -1.16). ANOVAs revealed a main effect (marginal by items) of ambiguity \( F_1(1, 47) = 4.14, p < .05; F_2(1, 35) = 3.90, p = .06 \), a significant main effect of disambiguation \( F_1(1, 47) = 474.45, p < .01; F_2(1, 35) = 956.23, p < .01 \) and a significant interaction between ambiguity and disambiguation \( F_1(1, 47) = 44.03, p < .01; F_2(1, 35) = 33.16, p < .01 \). This interaction was due to ambiguity in the ambiguous conditions as compared to the unambiguous controls. Finding ambiguity in this configuration of quantifiers provides prima facie evidence against Frey (1993)’s configurational account. Inverse scope readings seem to exist even in configurations where the second quantifier does not c-command a trace of the first quantifier. Two paired t-tests were computed as follow-up analyses In order to break down the interaction. Interestingly, these analyses revealed that the ambiguity affected both readings. It was not only the case that set diagrams disambiguating towards inverse scope were judged better in the ambiguous construction than in the disambiguated control condition \( t_1(47) = 5.29, p < .01; t_2(35) = 4.65, p < .01 \), but also the reverse was true for set diagrams only consistent with the strongly preferred linear interpretation. These were judged significantly worse in the ambiguous condition than in the appropriate control condition \( t_1(47) = -3.41, p < .01; t_2(35) = -3.43, p < .01 \).

In the os1 conditions the object quantifier exactly one preceded the distributive subject quantifier each. The potentially ambiguous construction was compatible with both scope readings, but the inverse reading was somewhat preferred over the linear interpretation (mean z-scores: 0.16 vs. -0.15). The scope unambiguous os1 control conditions patterned with the true and false fillers, respectively. They were judged true when combined with a picture only compatible with the linear interpretation and judged false when combined with an inverse picture. ANOVAs revealed significant main effects of ambiguity \( F_1(1, 47) = 35.0, p < .01; F_2(1, 35) = 52.1, p < .01 \) and of disambiguation \( F_1(1, 47) = 74.2, p < .01; F_2(1, 35) = 98.9, p < .01 \) and a significant interaction between ambiguity and disambiguation \( F_1(1, 47) = 109.4, p < .01; F_2(1, 35) = 197.2, p < .01 \). We computed a pairwise comparison between the linear and the inverse ambiguous conditions to test whether the preference for the inverse interpretation was statistically reliable. Paired t-tests revealed that the difference was marginal by participants and significant by items \( t(47) = 1.88, p = .07; t(35) = 2.58, p < .05 \). Thus, fully in line with Pafel (2005)’s prediction (cf. (15)) in this particular configuration of quantifiers inverse scope was slightly preferred over the linear reading.

The results of the os2 conditions are also fully in line with Pafel’s model. In the ambiguous construction linear scope was strongly preferred over inverse scope (mean z-scores: 0.40 vs. -0.60) but the difference between the two readings was less pronounced than in the disambiguated sentences (mean z-scores: 0.56 (linear) vs. -0.89 (inverse)). Thus, although strongly dispreferred, the
inverse reading is still possible in the ambiguous os2 construction. ANOVAs revealed a significant main effect of disambiguation ($F_1(1, 47) = 165.57, p < .01; F_2(1, 35) = 345.19, p < .01$) and a significant interaction between ambiguity and disambiguation ($F_1(1, 47) = 10.24, p < .01; F_2(1, 35) = 8.15, p < .01$) but the main effect of ambiguity was not significant ($F_{1/2} < 1$). The interaction is due to the fact that the difference between the judgments for the two scope readings was larger in the unambiguous cases than in the ambiguous ones. However, we were somewhat puzzled by the fact that the unambiguous os2 condition was rated better than the other two unambiguous conditions (i.e. os1 and so) in the inverse conditions. The explanation might lie in the pictures: It was only the inverse os2 conditions in which the extra line in the $\exists \forall$ disambiguations was critical. If participants overlooked the element which had two lines, they falsely responded “yes”. This demonstrates that it is important to keep an eye on the inherent complexity of the disambiguating picture. Crucially, however, this does not affect the interpretation of the interaction.

Comparing the relative differences between the linear and the inverse disambiguation in the so and in the os1 we see that Pafel’s predictions are not met in the so conditions. His theory would have predicted differences of the same size in both construction types. This was clearly not the case. While the linear and the inverse interpretation were more or less equally available in the ambiguous os1 conditions, the so conditions showed a strong preference for linear scope. So, may Frey’s account be still on the right track after all? As outlined in the introduction, Frey acknowledges the influence of non-syntactic factors such as lexical properties of quantifiers as well as intonation. Above, we have provided linguistic evidence that the tested quantifiers conform to Frey’s constrains on what should as quantificational expressions. It reamins an open question though whether participants sometimes implicitly chose an intonation that made inverse scope available (see, e.g., Fodor (2002) for empirical evidence for implicit prosody during silent reading). More specifically, Krifka (1998) extending Frey (1993)’s proposal by building on work by Büring (1997) has proposed that a rise-fall intonation contour might add the inverse reading, which is impossible otherwise. If this is correct, the ambiguity observed in the so conditions may thus be due to the fact that participants sometimes chose a rise-fall intonation contour in order to license inverse scope. Note that this explanation could also readily explain the difference between the results of the present experiment and Bott & Schlotterbeck (2012)’s results who did not find any evidence for inverse scope in the same construction in an online task presenting sentences in a self-paced reading task. It is plausible that in this kind of task readers are not able to construe the sentences with a rather marked rise-fall intonation contour. In an offline task as it was used here, however, choosing the appropriate implicit prosody is probably much easier. We therefore ran a pilot study explicitly addressing the role of intonation for scope inversion in subject-before-object sentences. The experiment was picture-verification experiment with auditory presentation of spoken stimuli controlled for intonation.
3.3 The Potential Influence of Intonation (Pilot Study)

We constructed 20 items in four conditions according to a $2 \times 2$ (intonation $\times$ disambiguation) within design. Target sentences with a universal, distributive subject quantifier preceding the object quantifier *genau ein* (exactly one) were embedded in short dialogues:

\begin{enumerate}
\item[(17)] a. A: Stimmt es, dass Andrea und Wolfgang viele gelbe Spielplättchen haben?
   B: Nein, JE/des Kind genau EIN\textbackslash gelbes Spielplättchen in seiner Spielecke.
   
   (A: Is it correct that Andrea and Wolfgang have many yellow pieces?)

   B: No, every child has exactly one yellow piece in its corner.)

b. A: Wieviele gelbe Spielplättchen haben Andrea und Wolfgang?
   B: Jedes Kind hat genau ein gelbes Spielplättchen in seiner Spielecke.
   
   (A: How many yellow pieces do Andrea and Wolfgang have?
   B: Every child has exactly one yellow piece.)
\end{enumerate}

Each dialogue was paired with two disambiguating pictures. Pictures used with the sample item are illustrated in Figures 7 and 8. The pictures always showed a table with partly overlapping play areas (‘corners’) of two children playing with geometrical pieces. Participants’ task was to decide whether B’s answer corresponds to a true description of the scenario. Speaker B’s answer in the dialogue (17-a) is an informative answer to the question under both scope readings. In its linear scope construal the answer is that none of the two children has many yellow pieces because each has only one (cf. Figure 7). Under the inverse scope construal, the answer says that there is only one yellow piece that Andrea and Wolfgang have in their joint area (cf. Figure 8), which again shouldn’t count as an instance of *many*. The dialogue (17-b) served as baseline control. Here, only the linear reading should be possible. This is also true in Pafel’s account with the first quantifier being the first, the subject and the distributive quantifier *jed-*. 

Dialogues were spoken and recorded by a male (speaker A) and a female speaker (speaker B). Speaker B was instructed to produce the answer in the rise-fall dialogues (17-a) with a rise-fall intonation. For the control condition (17-b) speaker B was asked to use what seemed the most appropriate intonation for the answer given the question. The produced intonation contours of speaker B were analyzed with respect to their fundamental frequency using the PRAAT...
software package (Boersma 2001). The average intonation contours in the two dialogue conditions are shown in Figure 9. In the control condition about half of the items were realized with two foci on the two quantifiers and the other half with verum focus on the auxiliary. When recording the stimuli, speaker B changed between these intonation contours and stated afterwards that to her both were possible answers to the how many question of (17-b).

20 native German participants judged the 20 items plus 60 filler dialogues in a latin square design making sure that they saw each item in only one condition. After inspecting the picture and listening to the auditorily presented dialogue they had to judge whether the answer of speaker B was a true description of the picture by providing a yes, true or no, false answer.

Acceptance rates were as follows. In the control condition the linear disambiguation was accepted 98% of the time, whereas pictures only consistent with the inverse interpretation were accepted 17% of the time. Acceptance rates were almost identical in the rise-fall conditions: linear scope disambiguations were accepted 96% of the time, whereas inverse scope disambiguations were accepted 20% of the time. This provides preliminary evidence against the view that doubly quantified German sentences with a subject-before-object word order can be more easily interpreted with inverse scope if realized with a rise-fall intonation. Coming back to the results of Experiment 1 we thus tentatively conclude that it is unlikely that participants implicitly imposed a rise-fall intonation to make inverse scope available. If this should turn out to be correct, the tested construction must in fact be considered scope ambiguous. We would like to emphasize though that the influence of intonation on scope interpretation needs to be investigated more carefully. We must leave this for future research.

Figure 9: Intonation contours of the target sentences in the pilot study used for dialogues of type (17-a) (left panel) and (17-b) (right panel). Fundamental frequency plots show mean F0-values split by sentence regions and 95% confidence intervals.
4 The Interplay of Scope Factors (Exp. 2 and 3)

The following two picture verification task experiments investigated the combined effects of *linear order*, *distributivity* and *discourse binding* in doubly quantified sentences. They address the third research question from above: do we find evidence for strictly additive influences of the different scope factors as predicted by multi-factor theories? Since answering this question may make it necessary to measure rather subtle differences between sentence conditions, we employed the Magnitude Estimation technique (Bard, et al. 1996) which has been widely applied in experimental syntax (e.g. Keller (2000), Featherston (2005) but see Weskott & Fanselow (2011), Sprouse (2011)).

4.1 Experimental Materials

The constructions tested in Experiment 2 and 3 are presented in (18) and (19). The complete list of items is provided in Appendix B.

(18)  a. Genau einen dieser Aufsätze hat jeder Student gelesen.
    Exactly one of these papers has each student read.
    Each student read exactly one of these papers.
  b. Genau einen dieser Aufsätze haben alle Studenten gelesen.
    Exactly one of these papers have all students read.
    All students read exactly one of these papers.
  c. Jeder Student hat genau einen dieser Aufsätze gelesen.
    Each student has exactly one of these papers read.
    Each student read exactly one of these papers.
  d. Alle Studenten haben genau einen dieser Aufsätze gelesen.
    All students have exactly one of these papers read.
    All students read exactly one of these papers.

In (18) the factors *linear order* and *distributivity* are manipulated. In (18-a) and (18-b) the $\exists!$-quantifier precedes $\forall$ whereas in (18-c) and (18-d) the order of quantifiers is reversed. The universal quantifier *jeder* (*each*) used in (18-a) and (18-c) is distributive whereas *alle* (*all*) in (18-b) and (18-d) is not. The factor *discourse-anaphoricity* is kept constant across conditions.

(19)  a. Genau einen Aufsatz haben alle diese Studenten gelesen.
    Exactly one paper have all of these students read.
    Each of these students read exactly one paper.
  b. Genau einen dieser Aufsätze haben alle Studenten gelesen.
    Exactly one of these papers have all students read.
    All students read exactly one of these papers.
  c. Genau einen Aufsatz hat jeder dieser Studenten gelesen.
    Exactly one paper has each of these students read.
    Each of these students read exactly one paper.
  d. Genau einen dieser Aufsätze hat jeder Student gelesen.
    Exactly one of these papers has each student read.
    Each student read exactly one of these papers.
In (19) the factors distributivity and discourse-anaphoricity are manipulated. While in (19-b) and (19-d) the first quantifier is related to a contextually salient set by using a partitive construction, in (19-a) and (19-c) it is the second quantifier that is linked to preceding discourse. In Experiment 3 word order was kept constant across conditions: the sentences always had object-before-subject word order.

4.2 Experiment 2: Manipulating Linear Order and Distributivity

In Experiment 2, we manipulated linear order (∀ before ∃ vs. ∃ before ∀) and distributivity (jeder vs. alle). A potentially scope ambiguous sentence like (18-a) to (18-d) was combined with a set diagram either disambiguating towards ∃!∀ or towards ∀∃!.

4.2.1 Methods

Participants: 56 native German speakers from Tübingen University (26 female, mean age 25.6 years) participated in the experiment for 5€. 5 additional participants were excluded from the analysis because of poor performance on the fillers.

Materials and Design: We used the 24 doubly quantified transitive sentences in the four variants in (18). The universally quantified phrase was the subject and the existentially quantified and discourse-bound exactly one of these-quantifier was the direct object. Each item was paired with two disambiguating set diagrams like the ones in Figures 1 and 2 resulting in a 2 x 2 x 2 factorial design with the within factors linear order, distributivity and disambiguation. Additionally, we prepared 36 distractor sentences using different quantifiers, negation and definite descriptions. 16 fillers were true and 20 were false. We constructed eight lists according to a latin square design.

Procedure Judgements were gathered using the Magnitude Estimation method (Bard et al. 1996). Participants judged sentence-picture pairs relative to a reference item. High values indicated that the sentence fits the picture, low values indicated a mismatch.

The experiment was a paper-and-pencil questionnaire. After reading written instructions, participants first completed a short practice session consisting of five trials. In the following experimental session, sentence-picture pairs were presented in an individually randomized order. Each sentence-picture pair was presented on an individual page in a small booklet with the reference item on top of each page. They were explicitly instructed to not turn the page to earlier parts of the experiment and complete the experiment in a page-ba-page fashion.

4.2.2 Results

The judgements of each participant were normalized by transforming them into z-scores. The true fillers were rated with a mean z-score of 0.76 (standard devi-
ation 0.80) and the false fillers with a mean z-score of -0.74 (standard deviation 0.74). Thus, participants paid attention to the semantic properties of the sentences.

The mean judgments are shown in Figure 10. Linear order and distributivity both showed effects in the expected direction. If a quantifier precedes the other it tends to take wide scope. Similarly, distributive quantifiers with the determiner jeder (each) take wide scope more easily than non-distributive alle (all). We computed repeated measures ANOVAs on the z-transformed judgments with the within factors distributivity (jeder vs. alle), order (∃! before ∀ vs. ∀ before ∃!) and disambiguation (∃!∀ vs. ∃∀). These revealed significant two-way interactions of distributivity and disambiguation ($F_1(1, 55) = 80.06; p < .01$; $F_2(1, 23) = 54.15; p < .01$) and order and disambiguation ($F_1(1, 55) = 20.71; p < .01$; $F_2(1, 23) = 41.73; p < .01$). Besides these interactions only the interaction between distributivity and order was significant ($F_1(1, 55) = 5.40; p < .05$; $F_2(1, 23) = 4.72; p < .05$). This was due to the fact that sentences with jeder before genau ein received higher ratings than the other constructions irrespective of the disambiguation. Crucially, the three-way interaction between order, distributivity and disambiguation was not significant ($F_1(1, 55) = 1.29; p = .26$; $F_2(1, 23) = 1.37; p = .26$). This shows that the prediction of purely additive effects of combined scope factors is in fact borne out.

Planned comparisons (with a Bonferroni corrected $\alpha$ of .0125) revealed significant differences between the $∃!∀$- and the $∀∃!$-disambiguation in the jeder before ein-construction ($t_1(55) = 5.36; p < .0125$; $t_2(23) = 6.12; p < .0125$) and in the ein before alle-construction ($t_1(55) = -5.21; p < .0125$; $t_2(23) = -6.59; p < .0125$) whereas the other two constructions showed no difference (all
Sentences with distributive jeder were fully compatible with both readings when genau ein preceded jeder but were preferred $\forall \exists$! when jeder was the first quantifier; the pattern was exactly the opposite for sentences containing the non-distributive determiner alle: alle before ein was ambiguous but ein before alle favored a surface scope interpretation.

The additivity of effects of linear order and distributivity is shown in Figure 11. The figure shows the observed difference scores between the mean judgments for the $\forall \exists$! disambiguation minus the mean judgments for the $\exists \forall$ disambiguation. The empirically observed difference scores are again plotted against the difference scores predicted by Pafel (2005)’s theory. The Figure shows that apart from minor exceptions there is a close correspondence between experimentally assessed scope judgments and scope values proposed by the theory.

4.2.3 Discussion

The results of Experiment 2 show that both linear order and distributivity have an effect on the scope of doubly quantified sentences. Quantifiers containing a distributive determiner like jeder (each) take wide scope more easily than quantifiers with a non-distributive determiner like alle (all). Similarly for linear order: linear scope is preferred over inverse readings. Interestingly, linear order and distributivity showed purely additive effects. The three-way interaction of order, distributivity and disambiguation was absent. This fits the predictions of multi-factor theories like Kuno (1991) and Pafel (2005) but without thresholds.
4.3 Experiment 3: Manipulating Distributivity and Discourse-Anaphoricity

Discourse-anaphoricity of the restrictor set has not received very much attention in the literature on quantifier scope. An exception is Musolino & Gualmini (2004) investigating inverse scope construals in children’s understanding of sentences including negation and quantifiers such as (20).

(20) a. The smurf didn’t catch two (of the) birds.
b. The troll didn’t find some of the jewels.

They found that discourse-anaphoric quantifiers significantly increased children’s ability to compute non-isomorphic, inverse interpretations, which were more or less absent in sentences without the partitive. Here, we are interested whether and how discourse related scope factors interact with lexical factors such as distributivity. In Experiment 4 we therefore manipulated distributivity and discourse-anaphoricity. Consider the sample item (21) repeated from (19).

(21) a. Genau einen Aufsatz haben alle diese Studenten gelesen.
   Exactly one paper have all of these students read.
   Each of these students read exactly one paper.
b. Genau einen dieser Aufsätze haben alle Studenten gelesen.
   Exactly one of these papers have all students read.
   All students read exactly one of these papers.
c. Genau einen Aufsatz hat jeder dieser Studenten gelesen.
   Exactly one paper has each of these students read.
   Each of these students read exactly one paper.
d. Genau einen dieser Aufsätze hat jeder Student gelesen.
   Exactly one of these papers has each student read.
   Each student read exactly one of these papers.

The universal quantifier in (21-a) and (21-b) is non-distributive; jeder in (21-c) and (21-d) is distributive. The partitive phrase dieser links one of the quantifiers to the context: in (21-a) and (21-c) it is the universal quantifier and in (21-b) and (21-d) it is the existential quantifier that is discourse-bound.

If these factors should be explicitly encoded in the hierarchical configuration we would expect asymmetrical dependencies. Manipulating the factor encoded in a higher position of the tree should block effects due to a manipulation of a factor encoded in some lower projection in the tree. By contrast, multi-factor theories assume no asymmetrical dependency but predict that factors should show purely additive effects. Experiment 3 tested these predictions.

4.3.1 Methods

The same sentence materials and fillers were used as in Experiment 3. We manipulated distributivity and discourse-anaphoricity but kept word order constant with exactly one always preceding the universal quantifier. The sentences were paired with the same disambiguating pictures as in Experiment 2 yielding a $2 \times 2 \times 2$ ($\text{distributivity} \times \text{discourse-anaphoricity} \times \text{disambiguation}$) within
The procedure was the same as in Experiment 2.

**Participants:** 24 native German speakers from Tübingen University (10 female, mean age 26.2 years) took part in the experiment. Each participant received 5€. Four additional subjects were excluded from the analysis due to poor performance on the fillers.

### 4.3.2 Results

The true distractors were rated with a mean $z$ score of 0.89 (standard deviation 0.68) and the false distractors with a mean $z$ score of -0.70 (standard deviation 0.80).

The distribution of scope readings in Experiment 3 is shown in Figure 12. We computed repeated measures ANOVAs analyzing $z$-transformed judgments with the within factors *distributivity (jeder vs. alle)*, *discourse-anaphoricity (discourse-bound $\exists!$ vs. discourse-bound $\forall$)* and *disambiguation ($\exists!\forall$ vs. $\forall\exists!$)*. Both *distributivity* and *discourse-anaphoricity* showed an influence on the distribution of scope readings. A discourse-bound quantifier tended to take wide scope more easily than a non-discourse-bound quantifier resulting in a significant interaction between *discourse-anaphoricity* and *disambiguation* ($F_1(1,23) = 7.33; p < .05$; $F_2(1,23) = 7.99; p < .05$). Furthermore, distributive *jeder* (*each*) had a stronger tendency for wide scope than non-distributive *alle* (*all*) leading to a significant interaction between *distributivity* and *disambiguation* ($F_1(1,23) = 17.00; p < .01$; $F_2(1,23) = 19.47; p < .01$). Beside these effects there was a main effect of *disambiguation* ($F_1(1,23) = 13.92; p < .01$; $F_2(1,23) = 32.55; p < .01$) which reflects a general preference for the $\exists!\forall$-interpretation. This preference
is due to constant word order in Experiment 2 with the existential quantifier always preceding the universal quantifier. The main effect of *distributivity* was also significant ($F_1(1, 23) = 13.48; p < .01$; $F_2(1, 23) = 13.05; p < .01$). This is due to the fact that the wide-scope universal disambiguations were judged better for *jeder* than for *alle* ($t_1(23) = 4.78; p < .01$; $t_2(23) = 4.88; p < .01$). Crucially, there was no three-way interaction between *discourse-anaphoricity*, *distributivity* and *disambiguation* ($F_1(1, 23) = .04; p = .85$; $F_2(1, 23) = .05; p = .83$). Thus, *discourse-anaphoricity* and *distributivity* showed purely additive effects.

Figure 13: Observed and predicted difference scores in Experiment 3. Both, observed and predicted mean difference scores in the four sentence conditions were normalized via z-transformations for better comparability.

The additivity of effects is also shown in Figure 13. The figure shows the observed difference scores between the mean judgments for the linear disambiguation minus the mean judgments for the inverse disambiguation. The empirically observed difference scores are again plotted against the difference scores predicted by Pafel (2005)’s theory. The Figure shows that there is again a rather close correspondence between experimentally assessed scope judgments and scope values proposed by the theory.

4.3.3 Discussion

This experiment provides evidence that *discourse-anaphoricity* also affects scope preferences: discourse-anaphoric quantifiers take wide scope more easily than non-anaphoric quantifiers. Again, the effects of the scope factors were purely additive, a result that corroborates the findings from Experiment 2. This pattern of effects is exactly what is predicted by multi-factor theories of quantifier scope.
5 General Discussion

The results of the reported picture verification task experiments showed that the scope relations in doubly quantified sentences depend on a number of factors: the linear order of the quantifiers, whether they are distributive and whether they are explicitly restricted by context. All three factors showed cumulative effects, that is, they added up rather than interacting with each other.

In section 1 we gave a systematic overview of theories on quantifier scope. Simplifying matters a great deal we focussed on two types of theories on scope interaction: configurational and multi-factor theories. Furthermore, because the presented experiments measured scope preferences in German, we picked out the configurational account by Frey (1993) and contrasted it with the multi-factor theory proposed in Pafel (2005). We now discuss the implications of the experimental findings for each of them in turn.

Configurational theories disambiguate the relative scope of quantifying expressions in the syntax. In Frey’s account inverse scope can only arise in configurations in which the second quantifier has syntactic command over a trace left in the base position of a quantifier in a fronted position. The results of Experiment 1 provide evidence against Frey’s scope principle. In the tested subject-before-object construction inverse scope was more acceptable than in the unambiguous baseline control condition even though quantifiers were chosen adopting Frey’s constraints for quantificational expressions. According to our criterion to determine scope ambiguity, we would therefore tentatively conclude that the answer to the first research question are German doubly quantified subject-before-object sentences scope unambiguous must be answered to the negative. Furthermore, we reported on a small pilot study suggesting that inverse scope in this construction type cannot be explained by what has been proposed with respect to prosodic influences on quantifier scope in German. We would like to point out that we found strong differences in scope preferences for sentences with fronted object quantifiers. As it stands, Frey’s theory does not make any specific predictions for these except for the general claim that they should in principle allow for both, linear and inverse scope construals. For these cases, the empirical coverage of the theory is thus rather limited and something needs to be added what should be responsible for these differences.

In the introduction we have mentioned configurational proposals that encode different scope factors in dedicated positions in the LF. In its original form (May 1977) the movement operation was fully unconstrained but more recent developments like Beghelli & Stowell (1997) make use of dedicated scope positions to account for individual properties of quantifiers. The cumulative nature of effects is, however, problematic for this type of theory because manipulating a factor that inhabits a higher scope position than another factor should rule out any influence of the latter. In Experiments 2 and 3 we did not find any evidence of asymmetrical dependencies between factors. It is unclear to us how this can be made compatible with such a configurational framework.

The findings of our experiments lend some support to multi-factor theories of quantifier scope. First, relative scope is influenced by a number of factors such as linear order, distributivity and discourse-anaphoricity. Second, scope
preferences seem to be graded rather than categorical. Third, they interact in a purely cumulative fashion. Contrary to Pafel (2005)’s theory, our findings do not support the assumption of thresholds distinguishing ambiguous from unambiguous sentences. All the constructions we tested in our experiments were scope ambiguous. Experiment 1 makes the strongest case for this claim. Even if one scope reading was extremely preferred over the other, the dispreferred reading was still judged better than a clearly unavailable reading in scope disambiguated controls. Moreover, the ratings of canonical subject-before-object sentences in Experiment 1 indicated that Pafel (2005)’s theory makes the wrong predictions for this kind of sentences. Here, the theory clearly has to be adjusted to fit the data.

Besides the theoretical implications, our study adds to the psycholinguistic work on quantifier interaction. Previous results were not sufficient to decide whether linear order has an influence on top of grammatical function. While VanLehn (1978), Fodor (1982), Gillen (1991), and Kurtzman & MacDonald (1993) provided support in favor of this claim, Ioup (1975), Catlin & Micham (1975), and Micham et al. (1980) argued that the grammatical functions of the quantifiers is more important than linear order. The constructions we used made it possible to manipulate linear order while keeping grammatical function constant. We found a strong effect of linear order indicating that the order of quantifiers influences scope independent of grammatical function. Furthermore, our study lends support to the findings of Tunstall (1998), Bott & Radó (2009), Radó & Bott (2012) and Brasovaeannu & Dotlacil (2015) in showing effects of distributivity. Finally, it adds discourse-anaphoricity to the list of scope factors, which has not been investigated before in the psycholinguistic research on scope preferences in adults but has been demonstrated to yield rather strong effects in childrens’ interpretation of relative scope of operators (Musolino & Gualmini 2004).

What’s more, our study is one of the first experimental studies on the combined influence of multiple scope factors. We have demonstrated that the systematic manipulation of factors claimed to be relevant for scope interpretation show the purely additive effects as predicted by multi-factor models that take into account the linear combinations of multiple constraints. In spite of being descriptively adequate, multi-factor theories are not very popular in the linguistic literature. There are reasons for this. Up to now, this type of theory just states that a factor is relevant for quantifier scope without offering an explanation why this should be the case. To also strive for explanatory adequacy, multi-factor theories need to address these questions.

References


A Items Tested in Exp. 1

(1) Jeden dieser Schüler lobte genau ein Lehrer voller Wohlwollen.

(2) Jeden dieser Vorfahren verehrte genau ein Adliger auf übertriebene Weise.

(3) Jeden dieser Angestellten beschützte genau ein Chef vor der drohenden Entlassung.

(4) Jeden dieser Patienten besuchte genau ein Therapeut letzte Woche zu Hause.

(5) Jeden dieser Spione informierte genau ein Geheimdienst über das drohende Attentat.
(6) Jeden dieser Anwälte erzürnte genau ein Mandant während des Gerichtsprozesses.

(7) Jeden dieser Lehrer grüßte genau ein Musterschüler auf dem Schulhof.

(8) Jeden dieser Kollegen mobbte genau ein Angestellter auf gemeine Weise.

(9) Jeden dieser Schützlinge betreute genau ein Erzieher besonders liebevoll.

(10) Jeden dieser Brüder enterbte genau eine Schwester ohne einen Moment zu zögern.

(11) Jeden dieser Studenten benachrichtigte genau ein Professor per E-Mail über die Note.

(12) Jeden dieser Cousins umarmte genau eine Cousine bei dem Familientreffen.

(13) Jeden dieser Behinderten begleitete genau ein Zivi am Donnerstag in die Kneipe.

(14) Jeden dieser Häftlinge beobachtete genau ein Wärter über die Videokamera.

(15) Jeden dieser Kunden beriet genau ein Banker mit viel Geduld am Telefon.

(16) Jeden dieser Zöglinge verwöhnte genau ein Pädagoge mit zu viel Aufmerksamkeit.

(17) Jeden dieser Ärzte verehrte genau ein Patient wie einen Gott.

(18) Jeden dieser Auftraggeber verärgerte genau ein Subunternehmer wegen später Lieferung.

(19) Jeden dieser Patienten beruhigte genau eine Krankenschwester während der Frühschicht.

(20) Jeden dieser Galleristen traf genau ein Maler in der Ausstellung.

(21) Jeden dieser Tänzer beschimpfte genau ein Choreograph nach der Premiere.

(22) Jeden dieser Entführer identifizierte genau eine Geißel bei der Gegenüberstellung.

(23) Jeden dieser Leibwächter entließ genau ein Popstar nach dem letzten Auftritt.

(24) Jeden dieser Trainer umarmte genau ein Fußballspieler nach dem Pokalendspiel.

(25) Jeden dieser Orchestermusiker kritisierte genau ein Dirigent wegen ungenauer Einsätze.

(26) Jede diese Nichten beschenkte genau ein Onkel mit einer Playstation.

(27) Jeden dieser Pfadfinder entdeckte genau ein Gruppenführer im Wald beim Zündeln.
Jeden dieser Filme finanzierte genau ein Produzent mit Werbeeinnahmen.

Jeden dieser Prüflinge nervte genau ein Prüfer durch fiese Fragen.

Jeden dieser Neffen unterstützte genau eine Tante durch reichliche Geldgeschenke.

Jeden dieser Nachhilfeschüler mochte genau ein Lehrer außerordentlich gern.

Jeden dieser Angehörigen ignorierte genau ein Familienmitglied auf dem letztjährigen Treffen.

Jeden dieser Betreuer imitierte genau ein Waisenkind mit großem Geschick.

Jeden dieser Touristen führte genau ein Reiseleiter auf den Ararat.

Jeden dieser Tanzpartner begeisterte genau eine Tänzerin durch einen einzigartigen Tanzstil.

Jeden dieser Köche belohnte genau ein Restaurantinhaber mit einer Extraprämie.

B Items Tested in Exp. 2/3

(1) Jede Oppositionspartei wollte genau einen dieser Gesetzentwürfe blockieren.
   'Each opposition party wanted to block exactly one of these law proposals.'

(2) Jede Expertengruppe hat genau einen dieser Sachmittelanträge vehement abgelehnt.
   'Each team of experts dismissed exactly one of these applications for material expenses vehemently.'

(3) Jede Person konnte genau einen dieser Verdächtigen eindeutig identifizieren.
   'Each person was able to identify exactly one of these suspects without any doubt.'

(4) Jede Familie hat genau einen dieser Zuschüsse beim Finanzamt beantragt.
   'Each family applied for exactly one of these grants at the revenue office.'

(5) Jede Studentin hat genau einen dieser Professoren über die Maßen angehimmelt.
   'Each female student adored exactly one of these professors exceedingly.'

(6) Jede Erzieherin hat genau einen dieser wilden Buben etwas zu hart bestraft.
   'Each nursery nurse punished exactly one of these wild boys slightly too heavily.'

(7) Jede Prüfungskandidatin hat genau einen dieser Theologiedozenten sehr gefürchtet.
   'Each examinee feared exactly one of these lecturers in theology.'
(8) Jede Zuschauerin hat genau einen dieser Fußballspieler lautstark angefeuert.
'Each female spectator loudly cheered for exactly one of these soccer players.'

(9) Jede Tanzschule hat genau einen dieser Volkstänze ins Programm aufgenommen.
'Each dancing school admitted exactly one of these folk dances to their program.'

(10) Jede Regierung hat genau einen dieser Verbrecher lange und intensiv gesucht.
'Each government intensively looked for exactly one of these criminals for a long time.'

(11) Jede Wählerin hat genau einen dieser Politiker als "blen Sexist beschimpft.
'Each female voter insulted exactly one of these politicians to be a bad sexist.'

(12) Jedes Labor hat genau einen dieser Stoffe für total unbedenklich erklärt.
'Each laboratory declared exactly one of these substances to be completely harmless.'

(13) Jede Weinhandlung hat genau einen dieser Rotweine aus dem Angebot genommen.
'Each wine seller took exactly one of these red wines out of its range of products.'

(14) Jede Bau-Firma hat genau einen dieser Wachdienste längerfristig engagiert.
'Each construction company hired exactly one of these security services for longer periods.'

(15) Jede Sopranistin hat genau einen dieser Operntexte außerordentlich gemocht.
'Each soprano liked exactly one of these libretti exceedingly.'

(16) Jede Gesellschaft hat genau einen dieser Arbeitsverträge als Vorlage genommen.
'Each company took exactly one of these labor contracts as a template.'

(17) Jedes Jurymitglied hat genau einen dieser epischen Texte überragend gefunden.
'Each member of the jury judged one of these epic texts to be outstanding.'

(18) Jede Buchhändlerin hat genau einen dieser Naturbildbände überaus gern empfohlen.
'Each book seller recommended exactly one of these illustrated books about nature very willingly.'

(19) Jede Journalistin hat genau einen dieser Filmschauspieler hartnäckig belagert.
'Each journalist stalked exactly one of these movie actors insistently.'
(20) Jede Klasse hat genau einen dieser Pilze überwiegend richtig bestimmt. 'Each class was largely able to correctly identify exactly one of these mushrooms.'

(21) Jede Reisegesellschaft hat genau einen dieser Riesenwasserfälle eifrig fotografiert. 'Each tour group took a picture of exactly one of these gigantic waterfalls.'

(22) Jede Psychotherapeutin hätte genau einen dieser Therapieansätze gerne ausprobiert. 'Each psychotherapist would have liked to test exactly one of these therapies.'

(23) Jedes Kind hat genau einen dieser Comics geradezu gierig verschlungen. 'Each child greedily consumed exactly one of these comics.'

(24) Jede Krankenschwester hat genau einen dieser Fremdsprachenkurse billiger bekommen. 'Each nurse got a special price for exactly one of these foreign language classes.'