The logic in language: How *all* quantifiers are alike, but *each* quantifier is different

CrossMark

Roman Feiman *, Jesse Snedeker

Department of Psychology, Harvard University, United States

ARTICLE INFO

ABSTRACT

Quantifier words like EACH, EVERY, ALL and THREE are among the most abstract words in language. Unlike nouns, verbs and adjectives, the meanings of quantifiers are not related to a referent out in the world. Rather, quantifiers specify what relationships hold between the sets of entities, events and properties denoted by other words. When two quantifiers are in the same clause, they create a systematic ambiguity. "*Every kid climbed a tree*" could mean that there was only one tree, climbed by all, or many different trees, one per climbing kid. In the present study, participants chose a picture to indicate their preferred reading of different ambiguous sentences – those containing EVERY, as well as the other three quantifiers. In Experiment 1, we found large systematic differences in preference, depending on the quantifier word. In Experiment 2, we then manipulated the choice of a particular reading of one sentence, and tested how this affected participants' reading preference on a subsequent target sentence. We found a priming effect for all quantifiers, but only when the prime and target sentences contained the same quantifier. For example, ALL-A sentences prime other ALL-A sentences, while EACH-A primes EACH-A, but sentences with EACH do not prime sentences with ALL or vice versa. In Experiment 3, we ask whether the lack of priming across quantifiers could be due to the two sentences sharing one fewer word. We find that changing the verb between the prime and target sentence does not reduce the priming effect. In Experiment 4, we discover one case where there is priming across quantifiers – when one number (e.g. THREE) is in the prime, and a different one (e.g. FOUR) is in the target. We discuss how these findings relate

* Corresponding author at: Department of Psychology, Harvard University, 33 Kirkland St, Cambridge, MA 02138, United States.

E-mail address: rfeiman@fas.harvard.edu (R. Feiman).

to linguistic theories of quantifier meaning and what they tell us about the division of labor between conceptual content and combinatorial semantics, as well as the mental representations of quantification and of the abstract logical structure of language.

1. Introduction

All languages have systematic rules of interpretation, allowing listeners to derive complex sentence meanings from combinations of words. The simplest hypothesis is that the rules that govern semantic and syntactic composition are perfectly coupled, so that every different meaning is expressed by a unique natural language sentence (Montague, 1970). Semantic ambiguities – cases where the same sentence has two different meanings – challenge this simple hypothesis, and thus are central to our understanding of the relationship between meaning and form. A particularly systematic type of ambiguity, known as a scope ambiguity, has been studied extensively by semanticists. Scope ambiguities arise in certain sentences that contain two quantifiers in the same clause (like EVERY and A). For example, take the sentence:

(1)   Every kid climbed a tree

This sentence could mean either that there is a single tree that every kid climbed, or that, for every kid, there is a (potentially different) tree that they climbed. The ambiguity in this sentence does not reflect an ambiguity in the speaker's mind, who surely knows which meaning she wants to convey. Thus we must have some format of representation that is less ambiguous than the English sentence and which precedes language production. In many semantic theories (see Heim & Kratzer, 1998; Hornstein, 1984; May, 1985, inter alia), this is captured by positing a level of representation, Logical Form (LF), separate from the surface form of the sentence, where the two interpretations are distinct. The two LFs for (1), ignoring tense, are presented below, with corresponding paraphrases:

(2a) $\forall x[Kid(x) \rightarrow \exists y[Tree(y) \wedge climbed(x,y)]]$
For every *x*, if *x* is a kid, then there exists a *y*, such that *y* is a tree and *x* climbed *y*

(2b) $\exists y[Tree(y) \wedge \forall x[Kid(x) \rightarrow climbed(x,y)]]$
There exists a *y*, such that *y* is a tree, and for all *x*, if *x* is a kid, then *x* climbed *y*

These LF representations are not ambiguous the way the English sentence (1) is because they specify the order in which the two quantifiers bind their variables. (2a) has the universal quantifier EVERY taking the widest possible scope. This is the interpretation with every kid climbing a potentially different tree. We refer to this as a "Universal-wide" or "U-wide" interpretation. (2b) is the interpretation where there is a single tree that every kid climbed. This sort of LF has the existential quantifier, A, taking wide scope, and we refer to it as an "Existential-wide" or "E-wide" interpretation.[1] Sentences like (1) are called 'scopally ambiguous' because the two possible meanings arise from the relative scope of the two quantifiers EVERY and A in LF.

---

[1] Note that the E-wide reading in (2b) entails the U-wide reading in (2a). We discuss this in Section 3.2.4.

The factors that affect how scope ambiguities are interpreted have been a topic of extensive study and debate within linguistics, but only recently has the application of psycholinguistic methods been used to ask about the *content* of LF representations (Raffray & Pickering, 2010; see also, Chemla & Bott, 2015). Raffray and Pickering were interested in how the construction of one LF representation affects the construction of another and what these patterns can tell us about the information that is relevant for constructing a logical form. Their study focused solely on sentences with EVERY and A. Participants read scopally ambiguous sentences like (1), and then picked, from two pictures, the one they thought best matched the sentence. On the prime trials, participants were forced to pick a picture corresponding to one scopal interpretation (e.g. the U-wide interpretation) because the other picture mismatched the sentence based either on the subject or object noun. On the target trials, which immediately followed the primes, participants read a new scopally ambiguous sentence, like (3) with the same two quantifiers, but different nouns.

(3)   Every hiker climbed a hill

Participants were then given a choice between two pictures, one for each of the scopal readings of the target sentence (see Fig. 3). Raffray and Pickering found that participants were more likely to pick a U-wide target picture after a U-wide prime than an E-wide prime. They interpret this priming effect as evidence of shared representational resources across the LFs constructed from the prime and target sentences.

Our experiments explore the locus of this priming effect, asking whether there are shared mental structures across sentences with different quantifiers and those with different verbs. We have two main goals. One is to home in on the nature of the operations used to construct different scopal interpretations, exploring which parts of a sentence these operations are sensitive to and which are irrelevant for these operations (i.e. what distinctions the operations abstract away from). For example, is it the case that the lexical entries of verbs are used in the process of assigning quantifiers to take scope relative to each other, or are individual verb meanings irrelevant to the scopal operations that are being primed in this procedure? The other, complimentary goal is to use our findings about the degree of abstraction in LF to speak to longstanding assumptions in the formal semantics literature about the distinction between conceptual content and the combinatorial properties of meaning, as well as to debates on whether different quantifiers with similar meanings (e.g., universals) assign scope using the same or different operations. Specifically we will explore: whether some quantifiers necessarily force a U-wide or E-wide reading (and therefore do not, in fact, give rise to scope ambiguities), whether some types of reading s are systematically more difficult to construct than others, and whether particular quantifiers (EVERY, ALL, and EACH; TWO, THREE and FOUR) share a common mechanism of scope assignment.

In the remainder of the introduction, we provide a brief outline of linguistic theories of quantifier scope, discuss how Raffray and Pickering's experiments bear on these theories, and outline our plan to exploit this paradigm to answer new questions.

1.1. Theoretical approaches to quantifier scope

How listeners derive an unambiguous representation from an ambiguous surface structure like (1) is a matter of longstanding debate (for a review see Ruys & Winter, 2011). Several classes of theories have been proposed. All posit additional operations on top of syntactic parsing of the surface structure, but they differ on whether these operations are syntactic (e.g. Quantifier Raising or QR, Chomsky, 1976; May, 1977, 1985; Quantifying-in, Montague, 1972; Rodman, 1976), semantic (e.g. Cooper storage, Cooper, 1983; Keller, 1988; type-shifting of expressions, Hendriks, 1988; type-shifting of composition rules, Barker, 2002) or involve a more complex mapping at the syntax–semantics interface (e.g., a lexicalized grammar where different quantifier words have different scopal mechanisms, Beghelli & Stowell, 1997; Steedman, 2012; See also Reinhart, 1997; Winter, 2001).

Evidence for and against different theories of quantifier scope has come from patterns of judgments about the availability of different readings for scopally ambiguous sentences. For example, prima facie
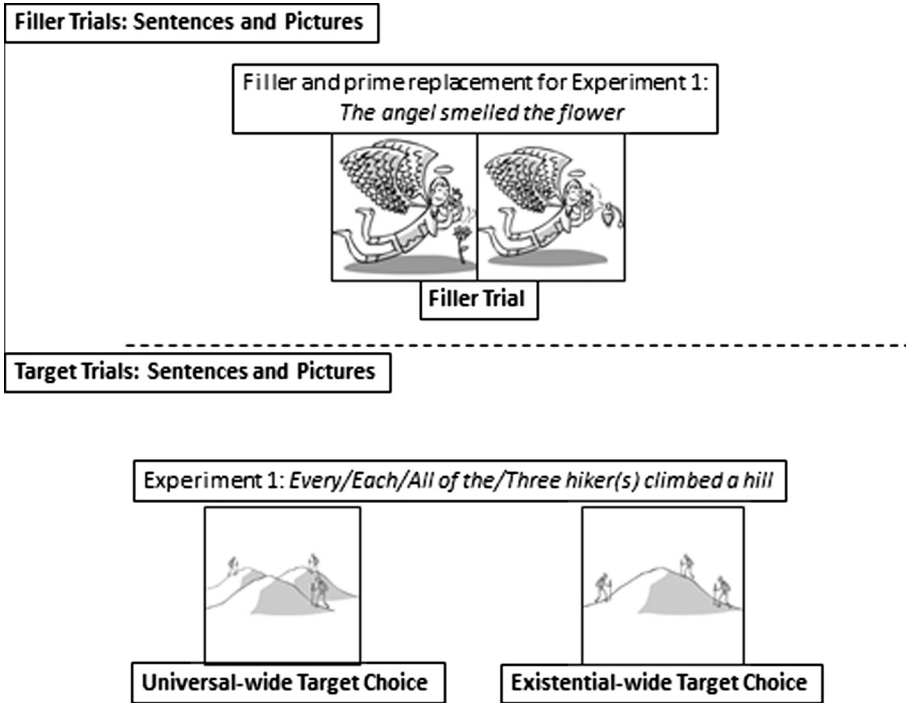
Fig. 1. Sample stimuli for Experiment 1. Filler trials (top) always directly preceded target trials (bottom). The correct picture choice on the filler trial depicted the corresponding reading of the sentence while the incorrect choice pictured a noun not present in the sentence.
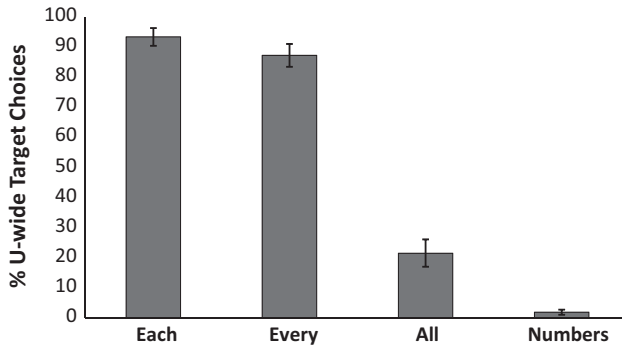

Fig. 2. Participants' rates of U-wide choices on ambiguous target trials in Experiment 1. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable.

evidence of parallel constraints on the movement of both Wh-phrases and quantifiers out of relative clauses provided some of the early evidence in favor of quantifier scope relying on syntactic operations like QR, but no theory yet exists that can easily account for all of the intricate patterns of preference and grammaticality judgments of different readings (see Ruys & Winter, 2011). Indeed, finding such a theory has been one of the key projects of semantics.
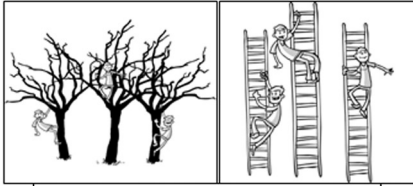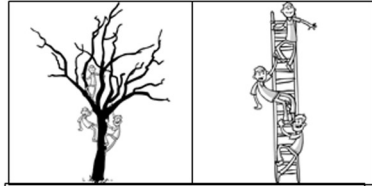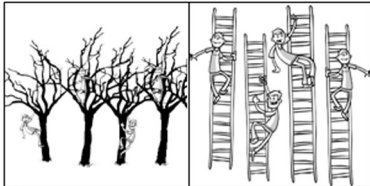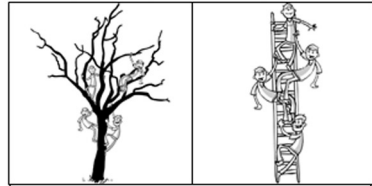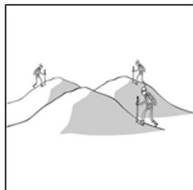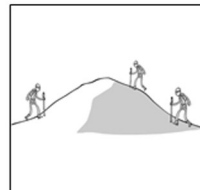
Fig. 3. Sample prime and target items for Experiments 2 and 4. Prime trials (top) always directly preceded target trials (bottom). There were two types of prime trials – U-wide and E-wide. The correct picture choice on the prime trial depicted the corresponding reading of the sentence while the incorrect choice pictured a noun not present in the sentence. Participants in the within-quantifier conditions of Experiment 2 saw prime and target sentences with the same quantifier words while participants in the between-quantifier conditions saw a different quantifier word in the prime than the target sentence. In Experiment 4, prime pictures were modified to show a different number of subject entities (e.g. extra kids) so that the prime sentence used a different bare numeral than the target (e.g. "Four kids climbed a tree" instead of three), while the target trial remained the same (e.g. "Three hikers climbed a hill").

It is worth noting that the term 'Logical Form' originally came from syntactic theories, though we will be using it in a theory-neutral manner. Following Hornstein (1984) and May (1985), LF is commonly described as a level of representation that is separate from surface syntax, directly interpreted by a semantic operation, but not directly connected to phonological form (with operations like QR therefore being *covert*, as opposed to *overt* syntactic operations like Wh-movement). Our use of the term LF throughout this paper does not indicate a commitment to this framework. All theories require some representations that disambiguate between the kinds of readings given by (2a) and (2b). We refer to these unambiguous representations as *LF*, and to the derivation of unambiguous meaning as the *construction of LF*, both for convenience and because the LF tradition has been the most dominant one in discussions of scope ambiguity. Importantly, however, the questions that we address in this paper arise in all theories of scope ambiguity, even though the terminology changes.

1.2. Accounting for scope preferences

Prior linguistic work has identified three factors that may influence which reading of a scope ambiguity is preferred. First, the order of the quantifiers in the sentence has been argued to be a major factor in their relative scope assignment, with quantifiers earlier in the sentence preferring wide scope (Johnson-Laird, 1969; Lakoff, 1971). In languages like English, the earlier quantifier is typically the quantifier that is higher in the syntactic tree. When these factors are disentangled, some theorists have argued that it is the higher quantifier (not the earlier one) that takes wide scope (Jackendoff, 1972; Lidz & Musolino, 2002; VanLehn, 1978). This observation is broadly consistent with syntactic theories (e.g. May, 1985), which posit that a costly raising operation is required when a quantifier that is lower in the syntactic tree takes wide scope, like when an object quantifier takes scope over a subject. If we assume that the parser avoids costly operations, this predicts a preference for linear scope readings (see Lidz & Musolino, 2002 for discussion).

Second, other theorists have suggested that scope preferences have a lexical component. In her seminal work, Ioup (1975) observed that different quantifier words seem to fall along a hierarchy based on the scope they prefer to take (see also Quine, 1960; VanLehn, 1978; Vendler, 1967 for related observations). This hierarchy is reproduced in (4), with quantifiers preferring wider scope appearing on the left:

(4)   EACH > EVERY > ALL > MOST > MANY > SEVERAL > SOME > A FEW

Ioup observes that, in general, quantifiers that select a larger set size seem to prefer wide scope over those that select smaller sets, but she makes no claims about why that should be. Since then, nearly every study of the comprehension and production of scopally ambiguous sentences has found lexically-based differences in the preferred scope assignments of quantifiers – though most of these studies have only examined sentences containing EVERY and A. The majority of them have found that EVERY prefers to take wide scope over A (generating the U-wide reading) independent of the linear or hierarchical order of the quantifiers (Bott & Radó, 2007; Catlin & Micham, 1975; Chemla & Bott, 2015; Clark & Kar, 2011; Filik, Paterson, & Liversedge, 2004; Gillen, 1991; Micham, Catlin, VanDerveer, & Loveland, 1980; Raffray & Pickering, 2010). The one exception to this generalization is a study by Kurtzman and MacDonald (1993), which found a slight preference for A to take wide scope over EVERY. This finding, however, has been argued to reflect factors specific to the dependent measure that they used (see Tunstall, 1998, for discussion). Thus the broader generalization appears to be secure.

However, it is not clear that the observed differences between sentences containing different quantifiers are caused by stable lexical properties that play a role in scope assignment. The third and final possibility is that scopal disambiguation is based on more general conceptual knowledge or expectations about the sorts of things that speakers are likely to say (Altmann & Steedman, 1988; Fodor, 1982; Katz, 1980; Van Berkum, Brown, & Hagoort, 1999). For example, if the context suggests that there might be a single referent for a singular indefinite (like A TREE) then the listener is more likely to get

an E-wide reading. But if the context suggests that multiple referents would be needed, then a U-wide reading would be preferred. On this hypothesis, differences in the number of things being discussed affect the plausibility of the message and so different quantifiers will show different patterns of interpretation, even though lexical knowledge plays no direct role in comprehension. For example, the observation that EVERY often takes scope over A could merely be a side effect of the fact that the discourses we encounter and construct with EVERY and A are more apt to support a U-wide interpretation than an E-wide one.

1.3. The abstractness of the operation

The present paper uses priming to explore the nature of the operations involved in disambiguating scope, following Raffray and Pickering (2010). These questions are separate from the debates between QR, type-shifting, and other linguistic theories, although they bear on those issues. Whatever theory one takes, there is a question about the abstractness of the disambiguating operations. For example, if the E-wide LF in (2b) is constructed by raising the object quantifier to a higher scope position in LF, then there is a question of how many raising operations there are and how they work – is there a single operation that raises the object quantifier or the second quantifier in a sentence, whatever it happens to be? Or is there a single operation for raising universal quantifiers but a separate operation for raising number quantifiers? Or are there different raising operations for each quantifier word? These same questions arise for a type-shifting account, only phrased in terms of the nature of the type-shifting operations. And perhaps some quantifiers are subject to movement by QR while others take wide scope through type-shifting or still another mechanism.

The factors that influence preferred scope readings bear on the issue of the abstractness of the scoping operations. Ioup's observations give us reason to suspect that there might be different operations for different quantifiers – the operation that acts on EACH and preferentially gives it wide scope might be different than the operation acting on A, which prefers to assign it narrow scope. That evidence is only suggestive, however. It could also be the case that there is a single abstract operation, but that quantifier words vary in the degree to which they call upon it. And as we mentioned above, it is possible that these differences reflect expectations about different situations in the world or the things others are likely to talk about, with different quantifier words being correlated with these expectations.

Raffray and Pickering's findings (2010) provide some initial insights into properties of the operations involved in constructing LF. Their first experiment, described above, demonstrates that scopal relations can be primed when the quantifiers are held constant but the nouns vary. In the subsequent experiments they established that the priming effect holds when the prime sentence is passivized, while the target stays active. The U-wide interpretation of a sentence like, *"A tree was climbed by every kid"* primes the U-wide interpretation of a sentence like "*Every hiker climbed a hill*". This pattern demonstrates that the primed operation (or representation) is sensitive to the underlying agent (the deep subject) but unaffected by changing the syntactic position of the arguments. These results, however, leave open the question of whether LF construction involves a number of different operations that are specific to individual quantifier words, like EVERY and A, or is one general operation that applies to many or even all quantifiers.

One way we address this question in the present work is by asking what role lexical content plays in LF construction. Are lexical items merely one cue (among many) that bias us toward particular scope relations? Or is LF construction differentiated according to the lexical or semantic content of quantifiers? This investigation also provides an opportunity for looking at the meaning shared between different quantifiers – if an LF is abstract relative to the universal quantifiers, but not others, it would suggest that the universals have an underlying representational similarity that plays a role in scope assignment and LF construction.

In Experiment 1, we begin with baseline measures of how scopally ambiguous sentences with different quantifiers are interpreted, when not preceded by any prime. We test the three English universals EVERY, EACH, and ALL, as well as numbers like THREE, FOUR and FIVE. This will allow us to test whether these different quantifiers show different scopal preferences, as suggested by Ioup (1975) and others. It will also provide a baseline for subsequent priming effects. In Experiment 2, we take the same

quantifiers and use them as both primes and targets in order to look at whether the priming effect found by Raffray and Pickering extends from one universal quantifier to another (e.g. EACH to ALL, EVERY to EACH, etc.), holding constant other factors like discourse, noun and verb content. We also examine whether there is a priming effect at LF between quantifiers from different families – from a universal in the prime to a number in the target, and vice versa. We find that priming seems to be limited to cases where the same quantifier appears in the prime and the target. Experiment 3 rules out the hypothesis that this pattern simply reflects a sudden drop off in priming due to the number of words that the prime and target have in common. We compare prime and target sentences that share the same quantifiers and have either the same verb or different verbs, and find robust priming effects in both cases. Experiment 4 looks at priming across sentences with different numbers. Although they are different lexical items, they are highly semantically related. We find priming that extends from one number to another.

2. Experiment 1: baselines

2.1. Method

2.1.1. Participants

For every condition in every experiment reported below, we recruited 128 unique participants on Amazon Mechanical Turk. Participants had to have had 98% of their previous work approved and had to be logging in from an IP address within the United States. Participants were barred from participating in more than one condition across all the experiments reported here. Participants were excluded for answering less than 90% of filler trials correctly in all experiments or if they indicated that English was not their first language. After these exclusions were applied, there were 116 participants in each of the EVERY (Mean age: 30.7, Range: 18–78), ALL (Mean age: 30.4, Range: 18–66) and number (Mean age: 34.8, Range: 19–73) conditions, and 119 in the EACH condition (Mean age: 30.4, Range: 18–57), who were included in the results below. The average participant in this and subsequent experiments indicated that they had some college education.

2.1.2. Stimuli

Experiment 1 used 24 target items, taken from Raffray and Pickering (2010, R&P hereafter), each of which consisted of a scopally ambiguous sentence (e.g. *"Every hiker climbed a hill"*) and two pictures. One picture depicted a U-wide reading (e.g. three hikers, each climbing a different hill), and the other depicted the E-wide reading (e.g. three hikers climbing the same hill). An example of a target trial is shown in Fig. 1. Depending on the condition, the subject quantifier in the sentence was either EVERY, EACH, ALL,[2] or one of the numbers (THREE, FOUR or FIVE). The object quantifier was always A.

As in R&P's Experiment 1, all of the sentences contained 12 past tense verbs (e.g., CLIMBED, CHASED, WATCHED) in an active construction, with animate subject nouns and either animate or inanimate objects in equal halves (e.g. *"Three hikers climbed a hill"* or *"Four lions chased a zookeeper"*). There were either three, four or five subject entities (with a third of the trials having each number). The number of object entities was either one (for the E-wide pictures), or the same as the number of subject entities (for the U-wide pictures; see Fig. 1). The side on which the U-wide and E-wide pictures appeared was counterbalanced within subjects.

Target trials were separated by three to five filler trials. There were 96 unambiguous filler trials in total. These, like target trials, consisted of a sentence and two pictures. However, fillers had unambiguous transitive sentences (e.g. *"The angel smelled the flower"*; see Fig. 1) which matched one of the pictures, while the other picture depicted a different subject or object. Of the fillers, 72 items were taken from R&P, altered only to make all sentences transitive, so that they would more closely resemble target items. The remaining 24 fillers were new items, constructed by reusing some of the picture

---

[2] In this, and all subsequent experiments, we used the partitive ALL OF THE instead of ALL to ensure that participants interpreted the sentence as referring solely to the depicted referents. *All kids climbed a tree* could be interpreted as a universal generalization about children as a kind (e.g., *All dogs descended from wolves*), and under this interpretation, the E-wide reading would be wildly implausible.

pairs from R&P's fillers, but changing the sentence so that the picture choice that was previously a distractor would be correct (e.g. *"The angel smelled the perfume"* for the same sample filler pictures in Fig. 1). To familiarize participants to the forced-choice task, an additional seven practice trials were provided. These had the same structure as the fillers. Stimuli were presented in one of two orders, one the reverse of the other.

2.1.3. Procedure

Participants first filled out a demographic questionnaire that asked, among other questions, whether English was their first language. They then completed the seven practice trials before proceeding with the experiment. On each trial, participants read a printed sentence centered above two pictures. Participants were instructed to select the picture that matched their interpretation of the sentence. They made their selection by clicking on one of the radio buttons located underneath each picture. Participants could not click on both buttons, and could not advance to the next trial until they clicked a button. Each trial was presented on a separate page, so that participants could not see multiple trials at once or return to a previous trial. In this and the other experiments, participants were aware that the experiment would time out after 1 h, encouraging them to complete it in one sitting. The experiment took an average of 12.8 min to finish.

2.2. Results and discussion

Our dependent variable was whether the participant selected the universal wide picture, which we will call a U-wide response, on the target trials. Results for each quantifier condition are presented in Fig. 2. The data was analyzed in the R programming language, v3.1.0 using the lme4 package, v1.1-8 (Bates, Maechler, Bolker, & Walker, 2015) to build a logit mixed-effects model (see Jaeger, 2008) with the maximal random effects structure appropriate for this experimental design (Barr, Levy, Scheepers, & Tily, 2013). In the common glmer syntax, the general form of the model specification for comparing between conditions was:

$$\text{TargetScope} \sim \text{TargetQuantifier} + (1|\text{Subject}) + (1|\text{Item}) + (0 + \text{TargetQuantifier}|\text{Item})$$

We did not include a random slope of Target Quantifier by subjects, since the manipulation of Target Quantifier was between-subjects. In our initial omnibus model we treated the ALL condition as the baseline and included three predictors, one for each of the remaining conditions (i.e., a dummy coding scheme). Table 1 shows the results of the statistical analyses and Fig. 2 shows the U-wide response rates by condition.

In this analysis, EACH and EVERY resulted in more U-wide responding than ALL, while the numbers resulted in less U-wide responding than ALL (see Table 1). Follow-up analyses contrasted pairs of conditions in descending order of U-wide response rate: EACH vs. EVERY ($z = 3.29$; $p = 0.001$; EVERY vs. ALL ($z = 15.1$, $p < 0.0001$); ALL vs. numbers ($z = -9.55$; $p < 0.0001$).

These findings strongly support earlier claims, grounded in linguistic intuitions, that different quantifier words influence scope resolution (Ioup, 1975; Quine, 1960; Vendler, 1967). At first glance these results appear to confirm Ioup's hierarchy, in which EACH has the strongest wide-scope preference, followed closely by EVERY, with ALL showing a preference for a narrow scope reading (i.e., an E-wide interpretation). But critically, Ioup's proposal is about the relative scope preferences between these different quantifiers (e.g., that EVERY will take scope over THREE). Our findings

Table 1
Logit mixed effects omnibus model for Experiment 1, with all quantifier baselines and the ALL condition as the dummy-coded reference level.

Predictor	Coefficient	Wald Z	p-value
Intercept	−2.75	−7.07	<0.0001
Condition = Every Baseline	6.85	14.34	<0.0001
Condition = Each Baseline	8.26	15.52	<0.0001
Condition = Number Baseline	−4.36	−7.81	<0.0001

demonstrate, instead, that there are differences in how likely each of these quantifiers is to take scope over another quantifier, the indefinite A.

Ioup observes that the ordering of quantifiers in her hierarchy can be characterized as a tendency for quantifiers picking out large sets to take scope over those picking out smaller sets. Ioup therefore predicts that universal quantifiers will consistently take scope over singular indefinites. This prediction is correct for EVERY and EACH but not for ALL, which was often given an E-wide interpretation, even though ALL picks out sets that are just as large, if not larger (see Tunstall, 1998) than the others. Similarly, on Ioup's proposal, numbers should take scope over A (since THREE clearly picks out more than one entity) but we instead find that the numbers had an extremely strong E-wide bias. In Section 6, we return to the implications of these findings for accounts that emphasize the influence of structural rather than lexical factors on scope preference.

We now proceed to the heart of the investigation, exploring whether and to what degree different quantifiers share common mechanisms of scope assignment by looking at patterns of priming. The results of Experiment 1 will provide a baseline for evaluating the priming effects of Experiment 2.

3. Experiment 2: between and within quantifiers

3.1. Method

3.1.1. Participants

An additional 128 unique participants were recruited for every prime–target quantifier pairing. The exclusion criteria were the same as for Experiment 1. After these exclusions were applied, the means of ages ranged between 29.4 and 33.1 across conditions, with comparably wide age ranges to Experiment 1. Table 2 shows the numbers of participants in each prime–target quantifier pairing after the application of the same exclusion criteria as in Experiment 1.

3.1.2. Stimuli and procedure

The stimuli of Experiment 2 were similar to Experiment 1. We used the same target trials, preceded by novel primes. The prime trials took the place of the extra fillers created for Experiment 1, and were placed immediately before the corresponding target trials. The target trials and the other fillers were unchanged from Experiment 1. Each prime sentence was accompanied by a prime picture that matched the sentence only on the reading that was being primed and a distractor picture, which mismatched the sentence on the subject noun (50%) or the object noun (50%), but matched the correct picture in terms of its quantifier-scope assignments (see for example, the U-wide and E-wide distractor items in Fig. 3). The stimuli (sentences and pictures for both primes and targets) were taken from R&P, modifying only the quantifier words in the primes and targets.

Across conditions, we manipulated the quantifier words used in the subject of the prime and target sentences (EVERY, EACH, ALL, or the numbers THREE, FOUR or FIVE) for a fully crossed 4 (Prime Quantifier) × 4 (Target Quantifier) between-participants design. Within participants, we manipulated Prime Scope (U-wide or E-wide).

We counterbalanced (within subjects and across items): the side of the correct prime picture, the side of the matching target picture, and whether the correct prime picture and matching target picture were on the same or opposite sides. The Prime Scope (U-wide or E-wide) was manipulated within item

Table 2
Numbers of participants in each prime–target quantifier pairing, after exclusion criteria were applied. Columns indicate the quantifier word in the target sentences and rows indicate the quantifier in the prime.

	Every	Each	All	Number
Every	112	112	108	116
Each	118	123	121	120
All	110	120	111	110
Number	108	118	117	110

and within participant by creating two counterbalanced lists. Two orders were generated as in Experiment 1, with the constraint that no more than two prime trials in a row had the same Prime Scope. In all other respects, the procedure was identical to Experiment 1. This experiment took 15.26 min to finish on average.

3.2. Results and discussion

Fig. 4 shows the results of Experiment 2, with conditions that share the same target quantifier grouped together. As in Experiment 1, the data was analyzed using logit mixed effects models. For all of the analyses described below, we modeled response type on the target trials (U-wide or E-wide), with random intercepts and random slopes relative to prime scope for subjects and items (uncorrelated with each other).[3] The model specification for looking at effects of Prime Scope within any given Within- or Between-Quantifier condition was:

$$\text{TargetAnswerScope} \sim \text{PrimeScope} + (1|\text{Subject}) + (0 + \text{PrimeScope}|\text{Subject}) + (1|\text{Item}) + (0 + \text{PrimeScope}|\text{Item})$$

For comparisons between multiple conditions, as when comparing effects of different prime quantifiers on the same target quantifier, the model specification is below:

$$\text{TargetAnswerScope} \sim \text{PrimeScope} * \text{PrimeQuantifier} + (1|\text{Subject}) + (0 + \text{PrimeScope}|\text{Subject}) + (1|\text{Item}) + (0 + \text{PrimeScope}|\text{Item}) + (0 + \text{PrimeQuantifier}|\text{Item})$$

Because the manipulation of Prime Quantifier was always between-subjects, we did not include a random slope of Prime Quantifier by Subject. The ease of collecting additional data on MTurk enabled us to replicate any borderline results. We replicate all conditions with *p*-values between 0.01 and 0.1 in the fixed effects of interest. We discuss these replications as we go, below.

Our central question is how different primes might influence the processing of the target sentences that follow them. From Experiment 1, we know that there are large baseline differences in preferred reading between target sentences with different quantifiers. Therefore, to look for priming effects, we hold the target sentence constant and look for changes in participants' responses as a factor of both the type of prime preceding any given target (U-wide or E-wide) and the different quantifier words in the prime sentences. Critically, this analysis strategy allows us to focus on responses to the exact same target sentence and explore whether they vary based on the preceding primes.[4]

3.2.1. Within-Quantifier priming

First, we want to establish whether the picture priming paradigm used here can indeed affect participants' target choice when the quantifiers in the prime and target sentences match. The use of EVERY in both prime and target replicates the design of Raffray and Pickering's Experiment 1 (and see Chemla & Bott, 2015), while the other conditions extend this within-quantifier priming design to the other quantifiers – EACH, ALL, and the numbers THREE, FOUR, and FIVE.

Looking at each Within-Quantifier condition separately, we find priming effects for EVERY ($z = 3.8$; $p = 0.0001$), EACH ($z = 4.89$; $p < 0.0001$), ALL ($z = 4.7$; $p < 0.0001$), and a marginally significant effect for the numbers ($z = 1.93$; $p = 0.05$). In a model including all four Within-Quantifier conditions, there were

---

[3] In almost all of the analyses in all experiments, the inclusion of a parameter of correlation between the random slope and intercept in the model (the default setting in lme4 version 1.1-7) resulted in overparameterization of the model. This is unsurprising given that Prime Scope – the variable by which the slope was being estimated – is binary. In no analysis did the inclusion of the correlation estimate improve model fit on an log-likelihood comparison test between models. This parameter was therefore excluded from all analyses.

[4] We tested all of the models reported here and in Experiment 3 and 4 with and without the two-level fixed effects of Order and List. We found some significant main effects and interactions of Prime Type and Prime Quantifier with these variables, reflecting the effects of specific items and orders. In no case did the inclusion or exclusion of these effects change the significance of the effects we were interested in. For simplicity and consistency, we therefore report the analyses without these fixed effects throughout.
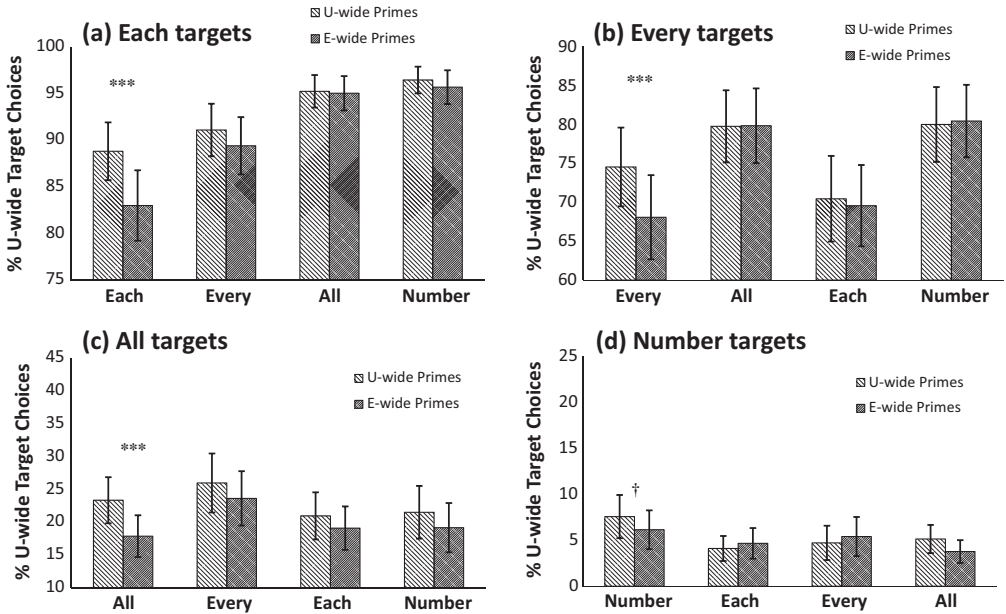
Fig. 4. Participants' rate of U-wide choices on ambiguous target trials in Experiments 2, separated by whether the target quantifier was (a) EACH (b) EVERY (c) ALL or (d) numbers. Note the different scales on the Y-axes between panels. Responses are broken down by whether the target trial was preceded by a U-wide or E-wide prime trial. The leftmost pair of bars in each graph shows the within-quantifier priming condition in that experiment, where the quantifier word is the same in both prime and target trials. The other pairs of bars show the between-quantifier conditions, with other quantifier words used in the prime sentences. All plots include original data only, with no data from replication runs. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable.

no significant interactions between Prime Scope and Condition, indicating that the size of the priming effects in each of the Within-Quantifier cases did not differ significantly from each other.

As the effect in the number-to-number priming condition was marginal at $p = 0.05$, it fit our criteria for replication. We replicated this condition, with 115 participants remaining after the usual exclusions. We found a significant priming effect in the replication ($z = 2.9$; $p = 0.003$) and no significant interaction of Prime Scope by Replication.

3.2.2. Between-Quantifier priming effects

Next, we want to examine whether there are significant priming effects between different quantifiers. To look for these effects, we performed four separate sets of analyses (one for each target quantifier), including in each set only the Between-Quantifier conditions that shared the same target. These analyses included the fixed effects of Prime Scope (U- or E-wide, coded as numeric 1 and 0 respectively) and Prime Quantifier (with three different prime quantifiers for every target quantifier, varying by target; for example, if EVERY was in the target, the three prime quantifiers included in this Between-Quantifier analysis would be EACH, ALL, and numbers), and the same random effects as the previous models. To look for a main effect of Prime Scope, we first compared the fit of a model with both Prime Scope and Prime Quantifier (but no interaction), to a model including only the effect of Prime Quantifier (using a Wald chi-square test). We found that including Prime Scope significantly improved model fit only in the case where ALL was the target quantifier ($\chi^2(1) = 5.92$, $p = 0.015$).

To look at the condition driving this Between-Quantifier effect, we constructed separate models for each Between-Quantifier condition where ALL was in target position (i.e. with different Prime Quantifiers). The only significant effect of Prime Scope was in the condition where the prime sentences contained numbers ($z = 2.1$, $p = 0.036$). Since this was a borderline effect that was not predicted by any

theoretical account, we replicated this condition with an additional 241 participants, after the usual exclusions. Including the data from both the original run and the replication, the effect was no longer significant in this condition ($z = -0.52$; $p = 0.6$). We therefore conclude that the original effect was a false positive. We conclude that there was no robust effect of Between-Quantifier priming across any two quantifiers.

Despite the lack of main effects of Prime Scope in the Between-Quantifier analyses for different target quantifiers, another planned comparison of theoretical interest is between the different types of Between-Quantifier cases. For example, if the quantifier in the target sentence is EVERY, primes with both EACH and a number constitute Between-Quantifier conditions. However, one of the central questions of this design is to ask whether there is more representational overlap between two quantifiers that may share a universal quantifier denotation (e.g. EVERY and EACH), as opposed to one universal quantifier and one existential with a specific set cardinality (e.g. EVERY and THREE). Looking only within the Between-Quantifier conditions, we separate universal primes EACH, EVERY, and ALL from non-universal number primes. Taking only the universal targets (EACH, EVERY, and ALL), we fit another logistic mixed effects model with fixed effects of Prime Scope and the between-subjects factor of Universal-vs-Number-Prime, as well as their interaction. If, for example, there was a priming effect of EACH on EVERY, but not of THREE on EVERY, we would expect to see a significant interaction term. However, there were no significant interactions for any of the universal target quantifiers: EACH, EVERY, or ALL.

Looking at models including both Prime Quantifier and Prime Scope, we also find significant main effects of Prime Quantifier in two cases: where EVERY is the target and EACH is the prime, and vice versa. These effects are not priming effects in the sense of being an influence of each prime trial on its subsequent target, but rather 'context' effects that range over the entire experiment, reflecting participants shifting their expectations about the scope bias of one quantifier based on repeatedly seeing that the other quantifier can have a normally dispreferred reading. The present experiment was not designed to look for such effects and is not ideally suited to test an effect of gradual listener adaptation. Across a number of replications, we found instability in these effects (with no similar instability in the priming effects). We provide an extensive discussion of these effects and their replicability in the Online Supplemental Materials.

3.2.3. Comparison of Between- vs. Within-Quantifier priming effects

Having found significant priming effects in all Within-Quantifier cases and no such effects Between-Quantifiers, we next test whether the effects in the two types of conditions differ from each other. As before, we analyze each target quantifier separately, this time comparing its Within-Quantifier condition to the three Between-Quantifier conditions (for example, the condition where EVERY is in both prime and target to the conditions where EVERY is in the target and other quantifiers are in the primes). We include fixed effects of Prime Scope (U- or E-wide), and Quantifier Overlap (Within- or Between-Quantifier) and their interaction as fixed effects, with the same random effects as the models above. We find that the interactions between Prime Scope and Quantifier Overlap are significant in all cases except for the numbers (where, recall, the Within-Quantifier priming effect was weakest, and where we did not include the replication data for the present analysis).

In sum, across the four target types, there was consistent scopal priming when and only when the prime and target shared the same quantifier. This pattern seems to suggest that the operations that construct logical form are in some way specific to the individual quantifiers that are being manipulated.

3.2.4. Addressing deflationary accounts of the priming effects

The pattern of findings reported above rules out several deflationary accounts of the locus of priming. First, that there are any priming effects at all rules out a possible speedy and shallow processing strategy that participants might have adopted on all but the target trials – simply avoiding the picture depicting nouns not mentioned in the sentence. Participants are not ignoring the quantifier content, especially given the difference between the presence of priming within-quantifiers, and the absence of priming between. Second, given that we use a picture-choice task for both primes and targets, one deflationary possibility is that any priming effect might be due to greater visual similarity

between a U-wide target picture and a U-wide prime – rather than an E-wide prime – picture. All U-wide pictures have pairs of agents acting on themes, while all E-wide pictures show many agents acting on a single theme. Similarly, the U-wide and E-wide pictures differ in the kinds of events they depict. In the U-wide picture, there are multiple events – each agent acts separately on a different theme. The E-wide picture depicts a single event, where all of the agents act together on the same theme. Another deflationary possibility is that the priming effect has little to do with linguistic representations, and occurs instead at the level of event representations.

However, the significant interaction between the Within- and Between-Quantifier conditions and the lack of priming Between-Quantifiers provide strong evidence that picture similarity and event similarity are not responsible for this form of priming. Not only are the pictures and event structures all the same in the Between-Quantifier conditions that do not prime as in the Within-Quantifier conditions that do, but the prime sentences in the Between-Quantifier conditions are ones that are capable of priming – they are the very same prime sentences that produce a priming effect in their respective Within-Quantifier conditions.[5]

This pattern of within- but not between-quantifier priming addresses another issue as well. For all sentences with a scope ambiguity between a universal and an existential quantifier, the E-wide reading always entails the U-wide reading. Thus, it might have been possible that participants seeing an E-wide picture would form, not an E-wide representation, but a U-wide representation in a special case (the case where every boy just happens to be climbing the same tree in sentence (1) and reading (2a)). But if participants were in fact forming a U-wide representation when seeing an E-wide picture, we would expect a *reduced* or nonexistent priming effect across the board – both types of primes would elicit a U-wide representation, and both target pictures would be consistent with it. Nor could priming be occurring for the specific representation of the coincidental case of the U-wide reading, since this would predict priming between, as well as within quantifiers. Not only does the pattern of priming in our data speak against this possibility, there are also theoretical reasons to think participants do not build a U-wide representation when seeing an E-wide picture. Mayr and Spector (2012) discuss the possibility that, since the E-wide reading entails the U-wide, but not vice versa, the U-wide representation may be enriched to contrast with the E-wide reading by excluding just the coincidental scenario that is compatible with both (e.g. *for every boy, there is a tree, but not the same tree*). This may be especially likely when the two types of readings are presented contrastively, as they are on each target trial.

Our findings demonstrate that scopal priming is mediated by linguistic representations: it depends on the identity of the quantifier and not just the picture or the event structure. Any deflationary account of priming would have to account for the distinction between the conditions where priming is observed (Within-Quantifier) and the conditions where it is absent (Between-Quantifier). One such account involves participants drawing a contrastive pragmatic inference in the Between-Quantifier conditions specifically, noting the similarity between the prime and target trials and reasoning that if a different quantifier is used in the target than the prime, then a different reading must have been intended. Note that on this account, priming would be present equally in all conditions, both Within- and Between-Quantifiers, with the contrastive effect serving to *suppress* priming in the Between-Quantifier cases. It would be a very strange coincidence indeed if the effect of anti-priming was close enough in size to the priming effect to combine for a non-significant effect in the Between-Quantifier cases. If participants are making a contrastive inference, this inference might be expected to differ if a prime–target pair contains two similar quantifiers, like EACH and EVERY, versus dissimilar ones, like EACH and THREE. Yet given the observed lack of priming in both cases, the contrastive suppressive effect would have to be close enough in size (or, even more coincidentally, vary in exact concord with a varying priming effect), so that it always cancels out any priming.

---

[5] Raffray and Pickering (2010, Experiment 4) also argue against an event- or picture-based interpretation of the priming effect, using a control condition that pairs the same pictures with generic sentences not containing quantifiers (e.g. "Kids climb trees"). They find no priming effect, but an overall inflated rate of target U-wide responding relative to sentences quantified with EVERY. This suggests participants may interpret such generics as closely matching U-wide pictures, and so the lack of priming may be specific to the U-wide reading of generics overriding any picture priming effect, rather than a demonstration that there is no such effect when the sentences are scopally ambiguous.

There is a related, but simpler account of the difference between Within-Quantifier priming and the lack of Between-Quantifier priming. The adjacent prime and target sentences in the Within-Quantifier cases are more similar to one another, at the phonological and lexical levels, than the corresponding sentences in the Between-Quantifier cases, simply because they share an additional word. Maybe this global increase in sentence similarity leads participants to draw comparisons between the two sentences or to infer that similar responses are expected. While the contrastive account above posits semantic priming across the board along with a suppressive effect in Between-Quantifier conditions, this account posits no semantic priming anywhere, with the appearance of priming Within-Quantifiers coming only as the result of participants taking a strategic approach to respond similarly to similar sentences. Experiment 3 explores whether reducing the similarity of the prime and target sentences, without changing the quantifier, will decrease or eliminate scopal priming. In Experiment 2 (and in R&P) the primes and targets always had the same verb. In Experiment 3, the primes and targets have different verbs (but the same quantifier). We chose to manipulate similarity by changing the verb because verbs define the event type and are thus likely to affect perceived similarity. Furthermore, while verbs have no privileged role in interpreting quantifiers, they do play a critical role in language processing. For example, in studies of syntactic priming (the priming of argument structure alternations), effects are often larger when two sentences share a verb than when they do not (see e.g., Pickering & Branigan, 1998). If scopal priming depends on mere similarity, it should be reduced or absent in this Between-Verb study. In contrast, if the effects in Experiment 2 are due to quantifier specificity in the operations that build Logical Form, then priming in this Between-Verb study should be equal to the corresponding condition of Experiment 2.

4. Experiment 3: between-verb priming

4.1. Method

4.1.1. Participants

An additional 128 participants were recruited via Amazon Mechanical Turk. The exclusion criteria were the same as for the previous experiments. After these exclusions were applied, 107 participants remained (*M* Age: 31.7, Range: 18–64).

4.1.2. Stimuli and procedure

Experiment 3 used the same stimuli and procedure as the Within-Quantifier condition with EVERY in both prime and target sentences, with the only difference being in the pairing of prime–target trials; prime–target pairs were scrambled so that target trials no longer followed prime trials containing the same verb (see Fig. 5 for an example). The average time to finish this experiment was 13.78 min.

4.2. Results and discussion

The data from target trials was analyzed using the same type of logistic mixed-effects model as Experiment 2, with the same random effects structure. We found a main effect of Prime Scope in this Between-Verb condition ($z = 3.11$; $p = 0.002$).

To compare the Between-Verb results from Experiment 3 to the equivalent Within-Verb condition of Experiment 2, another model included the EVERY-to-EVERY Within-Quantifier condition, including the same random effects, as well as the fixed effects of Prime Scope, Verb Overlap (Between- vs. Within-Verb) and their interaction. This model found no significant interaction ($z = -0.8$; $p = 0.42$). Wald chi-square tests on this model revealed no main effect of Verb Overlap ($\chi^2(1) = 0.03$; $p = 0.86$), but a highly significant main effect of Prime Type ($\chi^2(1) = 24.74$, $p < 0.0001$). Fig. 6 shows the rate of U-wide responding in Experiment 3 and the EVERY-to-EVERY condition of Experiment 2, which are similar. These results demonstrate that priming, at least from EVERY to EVERY, is not dependent on global similarity. Changing the verb between prime and target had no effect on the magnitude of scopal priming. In contrast, changing the quantifier in Experiment 2 eliminated priming.
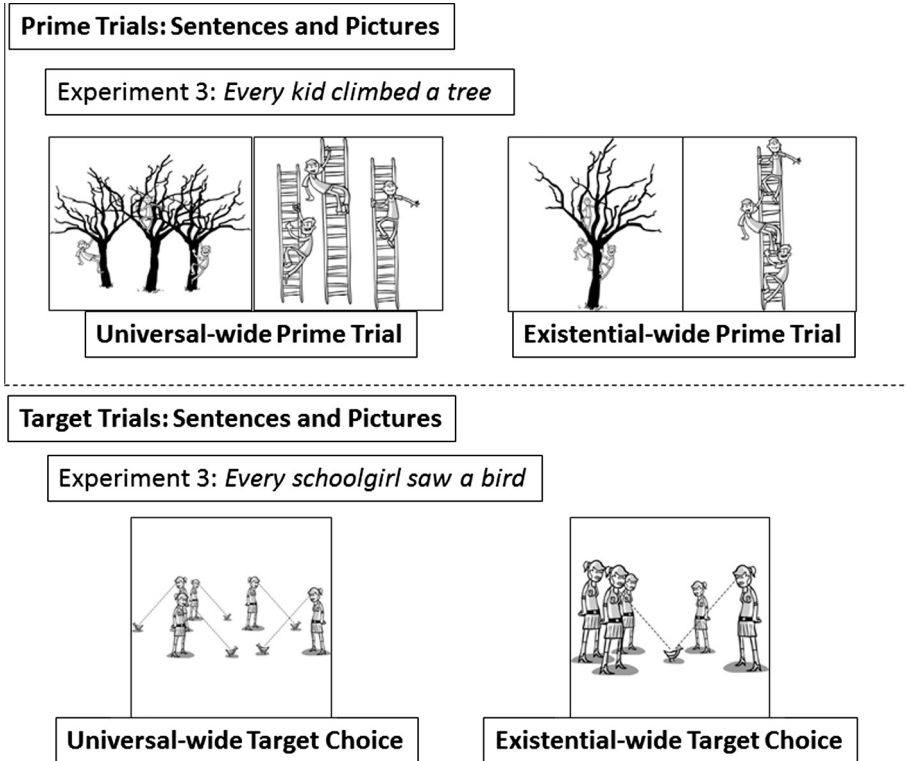
Fig. 5. Sample prime and target items for Experiment 3. Prime trials (top) always directly preceded target trials (bottom). There were two types of prime trials – U-wide and E-wide. The correct picture choice on the prime trial depicted the corresponding reading of the sentence while the incorrect choice pictured a noun not present in the sentence. The pairings of prime and target pairs was reshuffled in Experiment 3 relative to Experiment 2, so that prime and target sentences no longer shared their verb.

There are, however, two different ways in which priming could be quantifier-specific. First, priming could be tied to the phonological label of the quantifier. For example, each phonological label could have two semantic entries, one of which contains a type-shifting, QR, or some other operation to assign the quantifier wide scope, while the other does not. Thus priming of scope would occur only through a given phonological label. The second possibility is that quantifier specificity could be tied to substantive differences in the combinatorial properties of quantifiers. While scope ambiguity is a general property of quantification, some theorists have invoked distinct combinatorial features and mechanisms to explain both scopal and other semantic properties of EVERY, EACH and ALL (Beghelli & Stowell, 1997; Champollion, 2010; Steedman, 2012; see Section 6). Perhaps it is the differences between these meanings that are responsible for the specificity of the scopal priming effects.

To disentangle the effects of phonological form from shared meaning, in Experiment 4 we look for priming when a different number is used in the prime sentence than in the target sentence (THREE, FOUR or FIVE). Although there are many different analyses of the combinatorial semantics of number words (see, for example, Beghelli & Stowell, 1997; Heim & Kratzer, 1998; Steedman, 2012), within a given theory, different numbers always have identical combinatorial properties. Consistent with these theories, we adopt the assumption that the only differences between them are in the cardinality of the set they pick out. Thus, if different number words do not prime each other, it would suggest that the priming effect is mediated by their phonological labels. If sentences containing one number word do prime sentences containing a different one, it would show that priming can occur independent of the phonological label. That would in turn suggest that the lack of priming between quantifiers in Experiment 2
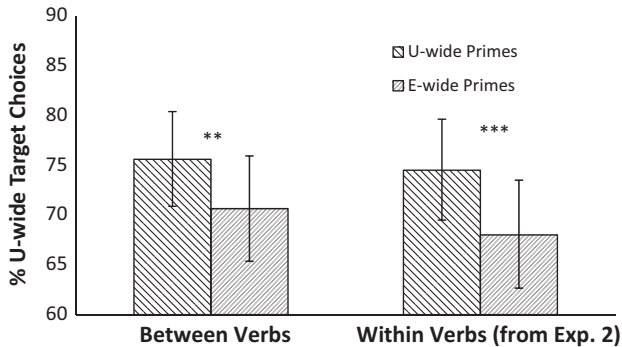
Fig. 6. Participants' rate of U-wide choices on ambiguous target trials in Experiment 3. Target trial responses are broken down by whether the target trial was preceded by a U-wide or E-wide prime trial. The right pair of bars shows the data from the Every-to-Every condition in Experiment 2. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable.

is a sign of differences in their meanings – specifically, in the semantic mechanisms of scope construction of each quantifier.

5. Experiment 4: between numbers

5.1. Method

5.1.1. Participants
An additional 128 participants were recruited. After applying the same exclusion criteria as in the preceding experiments, 112 participants (*M* Age: 28.5, Range: 18–64) were included in the final analysis.

5.1.2. Stimuli and procedure
The stimuli and procedure for Experiment 4 were identical to the Number-to-Number condition in Experiment 2, except that additional subject entities in the prime pictures were either added or removed from those pictures, so that the number of subject entities in the prime trial mismatched the number in the target trial (see Fig. 3). The target trials were unchanged for the sake of comparability with Experiment 2. For example, if the target trial presented a choice between two pictures – one with three hikers climbing one hill (E-wide), and one with three hiker–hill pairs (U-wide), the E-wide prime trial in Experiment 4 presented a choice between four kids climbing one tree or four kids climbing one ladder, and the U-wide prime trial presented a choice between four kid–tree pairs and four kid–ladder pairs. The prime sentence would correspondingly be, "Four kids climbed a tree", while the target sentence would remain as before, "Three hikers climbed a hill". The average time to finish this experiment was 14.92 min.

5.2. Results and discussion

Fig. 7 shows the data from the Between-Numbers condition compared to the number-to-number priming condition in Experiment 2. The data from target trials was analyzed using the same type of logistic mixed-effects model as the previous two experiments, with a maximal random effects structure. We again modeled Response Type on the target trials, with uncorrelated random slopes and intercepts for subjects and items, and with Prime Scope as a fixed factor. There was a significant effect of Prime Scope ($z = 2.07$; $p = 0.04$). In a second model including the (original runs of) Between-Number data from Experiment 4 and the Number-to-Number Within-Quantifier condition from Experiment 2 (where the number words used in a given prime–target pair were always the same), we added the fixed effect of Number Overlap and its interaction with Prime Scope. In this model, there
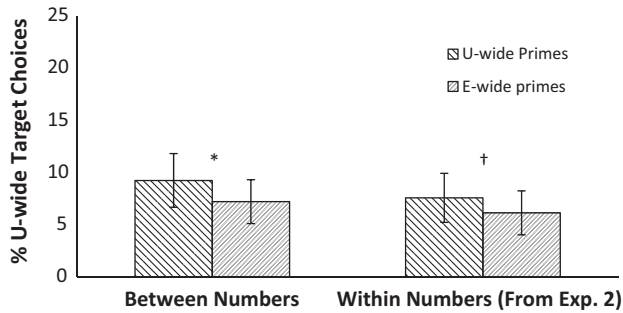
Fig. 7. The left pair of bars (Between-Numbers) shows participants' rate of U-wide choices on ambiguous target trials in Experiment 4, while the right pair shows participants' U-wide response rate on the Number-to-Number Within-Quantifier condition of Experiment 2 (Within-Numbers). Target trial responses are broken down by whether the target trial was preceded by a U-wide or E-wide prime trial. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable. This graph does not include data from the replication runs of either condition.

was no significant interaction between Number Overlap and Prime Scope ($z = 0.1$, $p = 0.92$). Wald chi-square tests on this model found no significant effect of Number Overlap ($\chi^2(1) = 0.39$; $p = 0.53$), but a significant main effect of Prime Scope ($\chi^2(1) = 9.94$; $p = 0.002$).

Since the effect of Prime Scope we found was significant at $p = 0.04$, it fit our criteria for replication. Replicating this condition, with 112 participants remaining after the normal exclusion criteria were applied, we found a significant effect of Prime Scope in the replication sample ($z = 2.73$, $p = 0.006$). Including both the original run and the replication data from both Experiment 4 and the Number-to-Number condition of Experiment 2, we again find a main effect of Prime Scope ($\chi^2(1) = 13.25$; $p = 0.0003$), no main effect of Number Overlap and no interaction of Number Overlap with Prime Scope.

Experiment 4 demonstrates that priming does not depend on whether two quantifiers share the same phonological label. Priming from THREE to FOUR or FIVE was as robust as priming from THREE to THREE. We discuss the significance of these findings in relation to linguistic theory in Section 6 (see Section 6.2).

6. General discussion

The four experiments described above shed light both on the mechanisms of scope ambiguity resolution, and on the representation of quantifiers. In Experiment 1 we find large baseline differences in scope-taking behavior between the quantifiers EVERY, EACH, ALL, and numbers. Experiment 2 uses a priming paradigm to examine how forcing the resolution of scope ambiguity in one sentence influences the interpretation of a subsequent one. We find that forcing a U-wide or E-wide resolution of the prime sentence influences how participants resolve a subsequent sentence, but only when the subject quantifier word in the prime and target sentences is the same. Changing the quantifier word eliminates priming. Experiment 3 shows that the difference in priming between the Within- and Between-Quantifier conditions of Experiment 2 cannot be due to overall similarity, since the priming effect is still present when the prime and target sentences have different verbs. Experiment 4 demonstrates that Between-Quantifier priming is possible when the quantifiers are different numbers, indicating that priming does not depend solely on the phonological label associated with the quantifier, but can occur when two different quantifiers have meanings that are sufficiently similar. These findings have implications for our understanding of the mechanisms behind quantifier-scope phenomena, for the question of whether and how the three universal quantifiers differ in their meanings, and for the interface between language and conceptual representations.

6.1. Lexical versus structural factors in preferred scope reading

Although many theorists have reported intuitive judgments about differences in scope preferences between quantifiers (e.g. Ioup, 1975; Quine, 1960; Vendler, 1967), the present studies provide

quantitative evidence for variation between the scope-taking preferences of different quantifiers, lending empirical support to a specific ordering of scope preference across the quantifiers (EACH > EVERY > ALL > numbers). The differences that we observed across the quantifiers in Experiment 1 are dramatic: participants chose the U-wide reading 93% of the time for EACH, but only 2% of the time for the numbers. Such stark differences are easily explained if quantifier scope depends largely on lexical factors, but they are difficult to reconcile with accounts on which the resolution of scope ambiguity is driven by structural factors such as linear order (Bunt, 1985; Fodor, 1982; Johnson-Laird, 1969; Kroch, 1974; Lakoff, 1971), surface syntactic c-command (Anderson, 2004; Jackendoff, 1972; Lidz & Musolino, 2002; May, 1985; VanLehn, 1978), or a thematic hierarchy in which quantified agents take scope over quantified themes (Grimshaw, 1990; Ioup, 1975; Jackendoff, 1972; Kurtzman & MacDonald, 1993). Our findings show that, to the extent that these other factors play a role (and they probably do; see Kurtzman & MacDonald, 1993) they can be overridden by the lexical or conceptual properties of individual quantifiers.

Why then have prior studies failed to observe these strong lexical biases? We suspect that it is because they have not looked for them. Most of the prior studies have focused on one pair of quantifiers at a time (mostly on EVERY and A; Catlin & Micham, 1975; Chemla & Bott, 2015; Clark & Kar, 2011; Filik et al., 2004; Gillen, 1991; Micham et al., 1980; Raffray & Pickering, 2010) and thus were not designed to compare the biases of different quantifiers. Given the many similarities between the different universal quantifiers – all are syntactic determiners, all specify that one set is the subset of another, all participate in scope ambiguities, and so on – it was not unreasonable to assume that quantifier-specific effects on preferred scope assignment would be minimal. Because most of the studies looking at structural factors did not employ the most strongly biased quantifiers (EACH and the numbers), they may have overestimated the effects of structural variables. Where lexical constraints are weak the effects of syntactic or thematic constraints may be clearest, but where lexical constraints are strong they may overwhelm these other biases.

Moreover, recent work by Chemla and Bott (2015) shows less support for structural factors even when looking only at the scopal relationship between EVERY and A. Using a picture priming paradigm, they find that assigning EVERY a wide scope reading in one sentence primes EVERY to take wide scope in a subsequent sentence, even when linear, thematic and syntactic structural factors are all stacked in the other direction. However, Paterson, Filik, and Liversedge (2008) did look at EACH-A scope ambiguities and found evidence of an interaction between quantifier identity and thematic hierarchy. Their approach was to examine reading time differences across structurally different sentences, indicating that the ability of EACH to override structural factors may at least in part depend on the paradigm as well.

That quantifier words vary so widely in their preferred scope readings suggests people store some information on how to construct combinatorial meaning in the lexical entries of different quantifiers. But what are these entries like? Our findings point to a distinction between two different levels of representation for quantifier meaning – the conceptual content of a quantifier word as one level and its combinatorial properties as another.

6.2. Combinatorial semantics versus conceptual content

An assumption commonly made in the linguistics literature (e.g. Heim & Kratzer, 1998) is that the combinatorial semantic system which produces complex sentence meanings from individual words is distinct from the conceptual system which stores each individual lexical item's content. This is why semantic denotations for many content words are conspicuously vacuous. For example, the denotation for SWIM typically looks something like this:

$$[\text{swims}] := \lambda x . x\text{SWIMS}$$

This denotation specifies the way in which the word "swims" combines with other words. SWIMS is a function that maps individuals to sets of individuals. Specifically, it maps entities (such as John, who might feature in the sentence, "John swims") to a set (in this case, the set of swimming things). But what constitutes a swimming thing? The combinatorial semantics do not say. The concept SWIM is invoked on the right hand side of the equation only as a pointer to some content, which is assumed to be stored separately from the combinatorial properties of the word (perhaps in a single amodal

conceptual store or perhaps as a distributed representation that is more tightly linked to perception and action). In just the same way, the differences in meaning between different number quantifiers are assumed to be stored in a separate conceptual system, with the semantic combinatorial system at LF providing pointers to the relevant conceptual entries. The combinatorial semantics of any number word specify that *There exists a set of M with cardinality N*, where the value of *M* is filled in by the noun being quantified, and the value of *N* is given by the number word. The conceptual content of any specific cardinality *N* – its intensional and extensional properties – would be stored in a separate conceptual inventory.

The pattern of priming effects that we find is quantifier-specific in a way that closely reflects this division of labor. That we find priming effects between different numbers (THREE, FOUR and FIVE), but not between different universal quantifiers (EACH, EVERY and ALL) supports the view that the former differ only in which concept they point to, while the latter differ in terms of their combinatorial semantic structure. The differences between the meanings of THREE, FOUR, and FIVE are handled by a different cognitive system than the ones responsible for the construction of LF and the resolution of scope ambiguity.[6] EACH, EVERY and ALL, on the other hand, do appear to differ in terms of their combinatorial properties.

While the division of labor between conceptual content and combinatorial properties is common to all linguistic theories of meaning, there is a wide array of specific theories of LF construction and scope assignment. We find that the mechanism that constructs LF is quantifier-specific rather than abstract, and that different quantifiers prefer very different scope assignments. While these findings provide new constraints on all theories of LF, they also fit more naturally with some accounts than with others.

6.3. A range of quantifier scope theories

For our purposes, theories of quantifier scope can be divided along a continuum, depending on how much they favor "lumping" or "splitting" of scoping mechanisms across different quantifiers. Theories on the lumping end posit a common mechanism responsible for all scope phenomena, while those on the splitting end hold that all quantifiers differ from each other in their scopal mechanisms. There are also many intermediate positions along the continuum, positing shared properties for some, but not all quantifiers. The extreme lumping end of the continuum, in particular, is densely populated. Theories at this end range from positing that all quantifiers share common syntactic operations, like Quantifier Raising, or QR (May, 1977, 1985) and quantifying-in (Montague, 1972; Rodman, 1976), to semantic mechanisms, like various forms of type-shifting; e.g. Barker, 2002; Hendriks, 1988) or Cooper Storage (Cooper, 1983; Keller, 1988).[7] All quantifier words in these theories are treated as Generalized Quantifiers in the sense first introduced by Montague (1972), with semantic denotations bearing the logical force of quantifiers in first-order predicate logic. Syntactically, too, all are members of the same category. In contrast, splitting theories posit different scoping mechanisms for some or all different quantifiers (e.g. Beghelli & Stowell, 1997; Champollion, 2010; Fodor & Sag, 1982; Reinhart, 1997; Steedman, 2012; VanLehn, 1978; Winter, 2001 and many others).

The combination of within-quantifier priming effects for all quantifiers and the lack of any priming effects between quantifiers cannot be explained by any single abstract operation acting on all of the different quantifiers during the online construction of meaning. It suggests instead that the operation or semantic feature responsible for scoping behavior is individualized by quantifier, possibly stored as part of each quantifier's lexical entry. Our findings are unexpected from the perspective of extreme lumping theories – the dominant framework in formal semantics for many decades – because, if the locus of priming was a single mechanism, it should have led to priming between all pairs of quantifiers.

We interpret this pattern as evidence that the procedures involved in constructing Logical Form during language comprehension are specific to individual quantifiers (or to narrow semantic classes

---

[6] See for example, Dehaene (1997), Le Corre and Carey (2007), for discussion of how the meanings of different number words are mapped to different representations in the Approximate Number System and other non-linguistic representational structures.

[7] See von Stechow (1991) for discussion of how semantic and syntactic approaches in this class become more similar as they posit more powerful operations on either end.

of quantifiers, such as numbers). For example, when we construct a representation where EVERY takes wide scope over A, we may activate a representational chunk that captures this precise scopal relation. If we later encounter a sentence which contains both EVERY and A, we will be more likely to activate the same scopal structure than the alternative. However, if we encounter a sentence with EACH and A, then neither of these representations will be relevant and no priming will occur.

6.4. Preferred versus possible scope readings

All of the semantic theories described above, regardless of their position on the lumping–splitting continuum, aim to account for the *possible* scope behavior of different quantifiers. They are largely mum on explaining differences in *preferred* readings. Even theories toward the splitting end, which posit semantic differences between quantifiers (e.g. Beghelli & Stowell, 1997), seek to explain which readings are allowed with which combinations of quantifiers, and not why some of the allowable readings are preferred to others. We find that all of the quantifiers we tested can take either wide or narrow scope relative to A and that even extremely dispreferred readings can become more available due to priming effects. Most linguistic theories of quantifier scope assume that differences in preferred readings are due to the effects of discourse context, and so foist off the explanation of these differences to the domains of pragmatics and domain-general, extra-linguistic cognition.

The results of Experiment 1 challenge this assumption. All of our target sentences were presented in exactly the same context regardless of the quantifier used: the filler trials, the instructions, the specific pictures presented, and the other words in the target sentences were identical across conditions. Only the quantifier in the subject noun phrase varied. Nevertheless, we found systematic baseline differences in the scopal preferences of the quantifiers we tested: while EVERY and EACH strongly preferred to take wide scope, ALL and the number quantifiers preferred to take narrow scope. In fact, every quantifier in our study was reliably different from each of the others. Since these differences cannot be chalked up to different pragmatic or discourse-level effects, they must be attributable to some lexically encoded difference between the quantifiers, based either on their semantic content or on their patterns of use.

One factor that is likely to play a role is the semantic feature of Quantifier Distributivity, which is posited to be an inherent part of the meaning of some quantifiers but optional or absent in others (see Beghelli & Stowell, 1997; Champollion, 2010; Roberts, 1987; Tunstall, 1998; Winter, 2001, among others). A distributive quantifier is one that forces each atom in the quantified noun set to have the predicated property applied to it separately. For example, EACH is argued to be necessarily distributive and thus unacceptable in contexts where the predicate is necessarily collective. Dowty (1987) argues that this accounts for the contrast between (7a) and (7b).

(7a)   *Each boy gathered in the yard

(7b)   All of the boys gathered in the yard

Dowty argues that (7a) is ungrammatical because GATHER requires a collective noun – a single thing cannot gather – and EACH can only provide individual atoms rather than collections. ALL, on the other hand, is able to take the entire set of boys as a collection (but see Champollion, 2010). This distinction between distributive and non-distributive quantifiers seems to pattern with the preferences we saw in scope assignment: EACH is typically argued to be mandatorily distributive and also showed the greatest preference for a U-wide scope reading; EVERY is often argued to be at least optionally distributive (Tunstall, 1998) and had a robust but less categorical U-wide preference; ALL, which seems to resist the distributive reading, showed a systematic E-wide preference. But it is critical to note that differences in distributivity alone cannot directly predict performance in this study. While the E-wide pictures in our study might seem more 'collective' than the U-wide pictures, both types of stimuli depict semantically distributive events. In both the U- and E-wide pictures, each subject is acting independently on its theme (e.g. each individual kid is climbing a tree in both types of pictures). Nevertheless, there may be a principled reason for why the E-wide picture could seem more in line with a collective

reading. Although both the U-wide and E-wide reading can be distributive, a collective reading severely constrains the availability of a U-wide reading. Take the sentence:

(8)   Every boy lifted a piano

If LIFTED receives a collective reading, where all of the boys lifted the same piano together, the only way for EVERY to take scope over A is in the coincidental case where for every boy, there is a piano they lift, but all the boys just happen to lift the same one (and this 'coincidental' reading may not be available if an E-wide reading is; see Mayr & Spector, 2012). The E-wide reading just has a single piano the boys acted on, and so is a natural fit for a collective interpretation. Although this observation leads us to speculate that there may be a principled link between quantifier distributivity and a preference for U-wide scope, we are not aware of any theoretical account of this kind.[8] We believe that investigating such a link is a promising target for future research in both semantic theory and psycholinguistics.

7. Conclusion

In relating our psycholinguistic investigation to semantic theories of quantifiers and quantifier scope, we attempt to connect the two fields to the benefit of both. In the field of semantics, theorists strive to provide adequate model-theoretic descriptions of the combinatorial machinery of language. In psycholinguistics and psychology, experimentalists strive to identify the mental representations that implement a combinatorial system in both language and thought. The experiments described above form a case study where these two projects constrain each other, providing important data for future theories to contend with.

We find empirical justification for a central assumption in linguistic theory – that representations of conceptual content are distinct from representations of the combinatorial properties of concepts. While for the universal quantifiers EACH, EVERY and ALL, the scope assigned to a quantifier in one sentence only influences the next sentence if it has the same quantifier, we also find that sentences with different numbers (THREE, FOUR and FIVE) influence each other. This suggests that the combinatorial machinery that assigns scope and constructs LF is sensitive to some differences between quantifiers (as between the different universals), but not to the differences between number concepts.

We also find large systematic differences across quantifiers with respect to their preferred scopal reading. These biases are not predicted by most leading theories of quantifier scope, which hold that structural, not lexical factors are primarily responsible for scope preference. Future linguistic theories should seek to explain the representational basis of these preferences, and how they relate to the mechanisms underlying quantifier scope assignment and LF construction. While we cannot neatly link these data patterns to existing semantic theories, we see these findings as part of a dialog between psycholinguistics and formal semantics which is still in its earliest stages. We are seeking out the phenomena that will drive future work, challenge our current theories and build a broader bridge between these fields.

Acknowledgments

---

[8] Beghelli and Stowell (1997) do link distributivity to the possible scope assignments of EACH and EVERY – they argue that these quantifiers can have a [+Dist] feature determine their landing site at LF. But they do not provide an account of why this would cause EACH and EVERY to *preferentially* take scope over A, since A can attach to a landing site either above or below DistP at LF. In other words, their account, like most others in the literature on scope, seeks to account for possible rather than preferred scope readings, and does not explicitly connect the two issues.

Science; and the scripting and piloting of participants in early versions of this work, contributed by Cara Aiello. Thanks also to extensive comments from Tom Urbach and from three anonymous reviewers.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cogpsych.2016.04.002.

References

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191–238.

Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity*. Unpublished doctoral dissertation. Evanston, IL: Northwestern University.

Barker, C. (2002). Continuations and the nature of quantification. *Natural Language Semantics, 10*(3), 211–242.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8. <http://CRAN.R-project.org/package=lme4>.

Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In A. Szcabolcsi (Ed.), *Ways of scope taking* (pp. 71–107). Netherlands: Springer.

Bott, O., & Radó, J. (2007). Quantifying quantifier scope: A cross-methodological comparison. *Roots: Linguistics in Search of Its Evidential Base, 96*, 53–74.

Bunt, H. (1985). *Mass terms and model-theoretic semantics*. Cambridge, UK: Cambridge University Press.

Catlin, J., & Micham, D. L. (1975). Semantic representations as procedures for verification. *Journal of Psycholinguistic Research, 4*, 209–225.

Champollion, L. (2010). *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Unpublished doctoral dissertation. Philadelphia, PA: University of Pennsylvania.

Chemla, E., & Bott, L. (2015). Using structural priming to study scopal representations and operations. *Linguistic Inquiry, 46*(1), 157–172.

Chomsky, N. (1976). Conditions on rules of grammar. *Linguistic Analysis, 2*, 303–351.

Clark, D. G., & Kar, J. (2011). Bias of quantifier scope interpretation is attenuated in normal aging and semantic dementia. *Journal of Neurolinguistics, 24*(4), 401–419.

Cooper, R. (1983). Quantification and syntactic theory. In D. Reidel (Ed.), *Synthese language library* (Vol. 21). Dordrecht, Netherlands.

Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.

Dowty, D. (1987). A note on collective predicates, distributive predicates, and 'all'. In F. Marshall (Ed.), *Proceedings of the third eastern states conference on linguistics (ESCOL 86)* (pp. 97–115). Columbus, OH: The Ohio State University.

Filik, R., Paterson, K. B., & Liversedge, S. P. (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 & E. Saarinen (Eds.), *Processes, beliefs, and questions: Essays on formal semantics of natural language and natural language processing* (pp. 129–164). Dordrecht: Reidel.

Fodor, J. D., & Sag, I. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy, 5*, 355–398.

Gillen, K. (1991). *The comprehension of doubly quantified sentences*. Unpublished doctoral dissertation. UK: University of Durham.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.

Hendriks, H. (1988). Type change in semantics: The scope of quantification and coordination. In E. Klein & J. van Benthem (Eds.), *Categories, polymorphism and unification* (pp. 96–119). Amsterdam: ITLI.

Hornstein, N. (1984). *Logic as grammar*. MIT Press.

Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball (Ed.). *Syntax and semantics* (Vol. 4, pp. 37–58). New York: Academic Press.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, Mass.: MIT Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.

Johnson-Laird, P. N. (1969). On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology, 21*, 1–13.

Katz, J. J. (1980). *Propositional structure and illocutionary force: A study of the contribution of sentence meaning to speech acts* (Vol. 6). Harvard University Press.

Keller, W. R. (1988). Nested Cooper storage: The proper treatment of quantification in ordinary noun phrases. In U. Reyle & C. Rohrer (Eds.), *Natural language parsing and linguistic theories* (pp. 432–447). Dordrecht: Reidel.

Kroch, A. (1974). *The semantics of scope in English*. Unpublished doctoral dissertation. Cambridge, MA: MIT.

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition, 48*, 243–279.

Lakoff, G. (1971). On generative semantics. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 232–296). Cambridge: Cambridge University Press.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395–438.

Lidz, J., & Musolino, J. (2002). Children's command of quantification. *Cognition, 84*(2), 113–154.

May, R. (1977). *The grammar of quantification*. Doctoral dissertation. Cambridge, MA: MIT.

May, R. (1985). *Logical form: It's structure and derivation*.Cambridge, MA: MIT Press.

Mayr, C., & Spector, B. (2012). Generalized scope economy – Not too strong! In M. Prinzhorn, V. Schmitt, & S. Zobel (Eds.), *Proceedings of Sinn and Bedeutung* (Vol. 14, pp. 305–321).

Micham, D. L., Catlin, J., VanDerveer, N. J., & Loveland, K. A. (1980). Lexical and structural cues to quantifier scope relations. *Journal of Psycholinguistic Research, 9*, 367–377.

Montague, R. (1972). The proper treatment of quantification in ordinary English. In J. Hinrlkka, J. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242).

Montague, R. (1970). English as formal language. In B. Visentini et al. (Eds.), *Linguaggi nella Società e nella Tecnica (Edizioni di Comunità)* (pp. 189–224).

Paterson, K. B., Filik, R., & Liversedge, S. P. (2008). Competition during the processing of quantifier scope ambiguities: Evidence from eye movements during reading. *The Quarterly Journal of Experimental Psychology, 61*(3), 459–473.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*(4), 633–651.

Quine, W. V. O. (1960). *Word and object*.Cambridge, MA: MIT Press.

Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension? *Psychological Science, 21*(8), 1090–1097.

Reinhart, T. (1997). Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy, 20*, 335–397.

Roberts, C. (1987). *Modal subordination, anaphora, and distributivity*. Doctoral dissertation. MA: University of Massachusetts at Amherst.

Rodman, R. (1976). Scope phenomena, 'movement transformations', and montague grammar. In B. H. Partee (Ed.), *Montague grammar* (pp. 165–176). New York: Academic Press.

Ruys, E., & Winter, Y. (2011). Scope ambiguities in formal syntax and semantics. In D. Gabbay & F. Gunthner (Eds.), *Handbook of philosophical logic* (2nd ed., pp. 159–225). Amsterdam: John Benjamins.

Steedman, M. (2012). *Taking scope: The natural semantics of quantifiers*.Cambridge, MA: MIT Press.

Tunstall, S. L. (1998). *The interpretation of quantifiers: Semantics and processing*. Unpublished doctoral dissertation. Amherst, MA: University of Massachusetts.

Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*(2), 147–182.

VanLehn, K. (1978). *Determining the scope of English quantifiers*. Unpublished Master's thesis. Cambridge, MA: MIT.

Vendler, Z. (1967). *Linguistics in philosophy*.Ithaca, NY: Cornell University Press.

von Stechow, A. (1991). Syntax and semantics. In A. von Stechow & D. Wunderlich (Eds.), *Semantik/semantics. An international handbook of contemporary research* (pp. 90–148). Berlin and New York: de Gruyrer.

Winter, Y. (2001). *Flexibility principles in Boolean semantics: The interpretation of coordination, plurality, and scope in natural language*.Boston, MA: MIT Press.