

The scalar inferences of strong scalar terms under negative quantifiers and constraints on the theory of alternatives*

Nicole Gotzner
ZAS

Jacopo Romoli
Ulster University

Abstract

Chemla & Spector (2011) have found experimental evidence that a universal sentence embedding a weak scalar term like *Every student read some of the books* has the strong inference that no student read all of the books, in addition to the weak one that not every student did (see also Clifton Jr & Dube 2010, Potts et al. 2015, Gotzner & Benz 2015). While it is controversial how this strong inference should be derived, there is consensus in the literature that this inference is an inference of the sentence above. On the other hand, the corresponding case of a negative quantifier embedding a strong scalar term like *No student read all of the books* with its corresponding potential strong inference that every student read some of the books, in addition to the weak one that some student read some of the books, is more controversial (Chemla 2009a,b,c, Romoli 2012, 2014, Trinh & Haida 2015). And, to our knowledge, there is no convincing experimental evidence for the existence of this strong inference. In this paper, we report on two experiments, building on Chemla & Spector 2011 and Chemla 2009c, systematically comparing sentences like the above with *every* and *no*. In our results, we find evidence for the strong inferences of both *every* and *no*. We discuss how standard theories of alternatives (e.g. Sauerland 2004b) can account for our data but also how they incur in an over- and under-generation problems which have been pointed out in connection with the combination of alternatives for sentences with multiple scalar terms (Fox 2007, Magri 2010, Chemla 2010, Romoli 2012). We discuss the two more constrained theories of alternatives by Fox 2007 and Romoli 2012 and we show that only the latter, combined with an independent account of the inferences of disjunction under universal quantifiers (Crnic et al. 2015, Bar-Lev & Fox 2016), can account for our data without incurring in the above-mentioned problems.

Keywords: Scalar implicatures, multiple scalar terms, embedded implicatures, localism-globalism debate

* For invaluable discussion and feedback, we thank Moshe Bar-Lev, Anton Benz, Emmanuel Chemla, Gennaro Chierchia, Danny Fox, Giorgio Magri, Paolo Santorio, Raj Singh; audiences at Xprag Chicago and Sinn und Bedeutung 21; our editor Daniel Rothschild, our reviewer Luka Crnic and two other anonymous reviewers for *Journal of Semantics*. Nicole Gotzner was supported by the Deutsche Forschungsgemeinschaft (DFG) (Grant Nr. BE 4348/4-1).

1 Introduction

Sentences containing more than one scalar item like (1-a) (*every* and *some*) have been used in recent years as an important testing ground for theories of scalar implicatures and theories of alternatives.¹ Such sentences have been probed experimentally, in particular Chemla & Spector (2011) have provided evidence that (1-a) gives rise to the strong inference in (1-b), in addition to the weak one in (1-c).

- (1) a. Every student read some of the books.
 b. \rightsquigarrow *No student read all of the books*
 c. \rightsquigarrow *Not every student read all of the books*

While it is controversial how the inference in (1-b) should be derived, there is consensus in the literature that the strong inference in (1-b) is an inference of (1-a). Moreover, experimental evidence for this reading exists from different methodologies (Clifton Jr & Dube 2010, Potts et al. 2015, Gotzner & Benz 2015).² The corresponding negative case of (1-a) in (2-b) is instead more controversial. Romoli (2012, 2014) has provided some argument suggesting that this inference exists as well (see also Trinh & Haida 2015). In line with this, Chemla (2009c) found no difference in comparing (1-b) and (2-b). As we discuss below, however, these data, while suggestive, are not conclusive evidence for the existence of (2-b). That is, to our knowledge, unlike the case of (1-b), the inference in (2-b) has not been demonstrated experimentally.³

- (2) a. No student read all of the books.
 b. \rightsquigarrow *Every student read some of the books*

1 See among others Fox 2007, Magri 2011, Chemla 2010, Romoli 2012, 2014, Chemla & Spector 2011, Trinh & Haida 2015, Geurts & van Tiel 2013, Sauerland 2010, Clifton Jr & Dube 2010, Benz & Gotzner 2014, Potts et al. 2015, Crnic et al. 2015; see also Geurts & van Tiel 2013 and Gotzner & Benz 2015 for an overview.

2 Note, however, that Geurts & Pouscoulous (2009) did not find evidence for such a reading and Geurts & van Tiel (2013) and van Tiel (2014) call into question whether we need scalar implicatures to account for Chemla & Spector's (2011) results; see Chemla & Spector (2015) for a response and section 4.3.2 for discussion of this in relation to our experiments.

3 Romoli's (2012) argument involved disjunctions like (i), which are constructed in such a way that the second disjunct does not entail the first one only if the strong inference in question exists (and it is computed locally; see Chierchia et al. (2012) among others for discussion of this type of sentences, generally called 'Hurford's disjunctions'). While suggestive, we find judgements about the acceptability of complex sentences like (i) too subtle to be satisfied with as the only argument for the strong inference above (see also Russell 2012 for discussion on this point).

- (i) None of my professors failed all of their students or Gennaro failed none and all of the others failed just some.

c. \rightsquigarrow *Some student read some of the books*

We think, however, in cases like (2-b) (and (1-b)), experimental evidence is particularly important, because, as argued by Sauerland (2004a) among others, it is difficult to assess intuitively whether a strong inference exists, in addition to a weaker one entailed by it. This is because the latter is of course always compatible with a situation in which the former is true, therefore it is difficult to disentangle introspective judgements about the weak inference versus the strong one.⁴

In addition, as we will discuss, it is controversial in the literature whether current theories of scalar implicatures and alternatives can derive the inference in (2-b). In particular, we summarise two studies below, Chemla (2009c) and Chemla (2009b), which investigate very different questions, but which are both based on the assumption that (2-b) could not be derived from (2-a) given current theories of scalar implicatures and alternatives. On the other hand, as we discuss, theories of alternatives of sentences with multiple scalar terms have been proposed, which do predict both the strong and weak inferences in (1-b) and (2-b)

In this paper, we report on two experiments systematically comparing sentences like (1-a) and (2-a), with the main goal of testing whether the inference in (2-b) exists. In our results we replicate Chemla & Spector's (2011) previous result finding evidence for the inference in (1-b). Crucially, we also find parallel evidence for the inference in (2-b). We show how the inferences above can be derived given standard theories of alternatives (e.g. Sauerland 2004b) and we discuss that such theories, however, incur in an over- and under-generation problems which have been pointed out in connection with the combination of alternatives for sentences with multiple scalar terms (Fox 2007, Magri 2010, Chemla 2010, Romoli 2012). We then turn to discuss the two more constrained theories of alternatives by Fox 2007 and Romoli 2012 and we show that only the latter, combined with an independent account of the inferences of disjunction under universal quantifiers (Crnic et al. 2015, Bar-Lev & Fox 2016), can account for our data without incurring in the above-mentioned problems.

The paper is organized as follows: in section 2, we review three previous studies, which have looked at sentences with *every* embedding *some* or *no* embedding *all*, and we discuss how there is still no convincing experimental evidence for the strong inference of the latter. This constitutes the main motivation of the two experiments that we report in sections 3 and 4, along side with a discussion of methodological issues related to this type of experiments. In section 5, we discuss our results and how they relate to theories of alternatives with multiple scalar terms and theories of scalar implicatures. In particular, we discuss how deriving the inferences above interact

⁴ See also Chemla & Spector (2011) and Meyer & Sauerland (2009) among others and section 4.3.2 for discussion on this point.

with over- and under-generation problems, which have been raised in connection with sentences with multiple scalar terms, and with the debate on the existence of embedded scalar implicatures. We conclude the paper in section 6.

2 Previous Studies

We summarise three studies on which our experiments build on. First, we discuss the result by Chemla 2009b about scalar terms vs presuppositional triggers in the scope of *no*. Second, we review Chemla’s (2009c) study, comparing scalar terms in the scope of *every* vs *no*. Finally, we summarise the results of Chemla & Spector (2011) about the implicature of *some* in the scope of *every*. These three studies are particularly relevant for our purposes for two reasons. First, as mentioned above, the study by Chemla (2009b) and that by Chemla (2009c), are both based on the assumption that the strong inference of sentences with *no* embedding *all* cannot be derived by current theories of scalar implicatures and alternatives. Second, the results of Chemla & Spector (2011) and Chemla (2009c), when taken together, constitute suggestive, albeit not conclusive, evidence that this inference of negative quantifier sentences embedding strong scalar terms should instead be derived. This two aspects of these previous studies are the basis of the main motivation of the two experiments that we report in the subsequent sections.

2.1 Scalar terms versus presuppositions

Chemla (2009a,b) compares (the corresponding French sentences of) sentences like (2-a), with its potential strong and weak inferences in (2-b) and (2-c), and sentences like (3-a), where *no* embeds a presuppositional trigger, *his*, and the corresponding potential strong and weak inferences in (3-b) and (3-c). Sentences like (3-a) are extensively discussed in the literature on presupposition projection and, in the same way as in the scalar implicatures literature, it is controversial whether (3-a) is associated with an inference like (3-b), one like (3-c), or possibly both.⁵

- (3) a. No student takes good care of his computer.
 b. \rightsquigarrow *Every student has a computer*
 c. \rightsquigarrow *Some student has a computer*

In addition to investigating what possible inference a sentence like (3-a) could have, Chemla (2009c) compares sentences like (3-a) and (2-a) with the goal of testing whether scalar implicatures and presuppositions should be treated alike, as proposed

⁵ See Heim 1982, Beaver 1994, Chierchia 1995, Beaver 2001, Charlow 2009, Schlenker 2009, Rothschild 2011, Romoli 2012, 2014, Sudo et al. 2011, Sudo 2012, Fox 2012, Zehr et al. 2015 among many others.

in the literature by himself (Chemla 2010) and others (Abusch 2010, Simons 2001, Romoli 2012, 2014). This aspect of Chemla's study is based on the assumption that a scalar implicature theory cannot derive the strong inference in (2-b) and thereby *qua* theory of presuppositions would not be able to derive the inference in (3-b).⁶

The study was an inferential task: participants were presented with sentences like (2-a) and asked how much they would infer (2-b) or (2-c) and similar for (3-a) versus (3-b) or (3-c). In his results, Chemla finds an effect of type of trigger (scalar vs presuppositional), with the presuppositional inferences being endorsed more overall, and an interaction between the strong and weak inferences and type of trigger. That is, the results show that while both inferences in (3-b) and (3-c) are equally strongly endorsed, the inference in (2-c) is endorsed much more than that in (2-b). In addition, participants endorsed the inference in (2-b) only 25% of the time.

In sum, Chemla (2009a,b) finds evidence for the inference in (3-b) but no evidence for that in (2-b). Moreover, his results suggest that if the inference exists, it is certainly much weaker than that of the corresponding negative quantified sentences embedding a presuppositional trigger. Of course, while these results contribute no evidence for the strong inference of *no* embedding *all*, they do not show that this inference doesn't exist either.⁷ In the next section, we turn to another study by Emmanuel Chemla again, looking at the case of *every* embedding *some* versus that of *no* embedding *all*, in comparison.

2.2 *Every versus No*

Chemla (2009c) is also an inferential task investigating in comparison cases like (1-a) and (2-a) and their strong inferences in (1-b) and (2-b). In this study, participants were presented with such sentences and their candidate strong inferences and the task was to indicate whether the sentence suggested the corresponding inference. Chemla also compared the cases in (1-a) and (2-a) to their non-quantified counterparts (e.g. *John read some of the books* and *John didn't read all of the books*).

One of the main goals of the study was to compare the predictions of theories allowing embedded derivations of scalar implicatures (local theories) versus theories that do not (global theories); that is, to test the prediction of a theory that would allow

⁶ See in particular the discussion in section 1.3 of Chemla 2009c; see also the discussion in section 2.2. and Appendix C of Chemla 2009a.

⁷ Moreover, even if we insist on taking presuppositions as a type of scalar implicature, the difference between the strength of the inference in (3-b) versus that in (2-b) is also not that surprising, considering in particular results like that in Van Tiel et al. 2016, showing how the strength of scalar inferences varies considerably. Nonetheless, the challenge for a scalar implicature theory of presuppositions remains: the difference between the case in (3-a) and (2-a) has to be accounted for, if *all* and *his* are treated essentially alike. See Chemla 2010, Romoli 2012, 2014 for different responses to these data from the perspective of a scalar implicature theory of presuppositions.

local scalar implicatures versus one that would not.⁸ In particular, Chemla (2009c) assumes that deriving strong inferences in either case globally is not possible and concludes that the comparison between *every* and *no* sentences should either reveal no strong inference at all or only a strong inference for *every* (derived locally).⁹ The former case would have favored a globalist account, while the latter a localist one.

While, as we discuss below, it cannot help us addressing the local-global debate, the set-up of Chemla's experiment can potentially give us insight with respect to the question as to whether the strong inference of *no* embedding *all* exists at all. And indeed Chemla (2009c) found that while the weak inference was accepted to a greater extent than the strong inference, there was no difference between the cases of *every* and *no*. Nonetheless, while Chemla's results is suggestive evidence for the existence of the inference in (2-b), we think that it is not enough to settle the issue. There were two things in particular which were missing in Chemla's study. Firstly, the weak and the strong inferences of the quantified cases were not compared between each other. Secondly and more importantly, there were no baseline controls of a no-inference case. For these reasons, we cannot tell whether participants actually computed the strong inference (for either *every* or *no*) or whether they found the inference just compatible with the sentence. Therefore the non-difference Chemla finds between the two quantifiers cannot tell us whether either of the two strong inference exist.

We turn now to discuss how Chemla & Spector (2011) improves on this score with respect to the case of *every*.

2.3 Evidence for the strong inference of *every* embedding *some*

In their study, Chemla & Spector (2011) used a picture verification task, rather than an inferential one, to test the inferences weak scalar terms give rise to under embedding. In their first experiment, participants were presented with sentences like (1-a), paired with a picture in which certain readings of the sentence were true. Participants' task was to indicate to what extent the sentence was true in the situation represented by the picture (by adjusting a slider ranging from 0 to 100). Chemla and Spector compared participants' truth value judgements across four experimental conditions: a FALSE condition in which neither reading of the sentence was true, a LITERAL condition in which only the literal reading was satisfied, a WEAK condition in which the weak and literal reading were true, and a STRONG condition in which all readings were true. The main finding was a graded pattern of responses such that

⁸ See section 5.2 below for discussion of the global vs local debate in the scalar implicatures literature.

⁹ That is, Chemla (2009c) is aware of the derivation of (2-b) (and (1-b)) globally by making use of more alternatives, which we discuss below, and discusses this possibility in the Appendix. However, he dismisses it precisely because of theories of alternatives like Fox 2007 that would block such derivation. We will come back to this in the general discussion below.

participants gave highest ratings in the strong condition that satisfied all readings, followed by the weak, literal and false conditions. Chemla and Spector interpreted these results as indicating that the strong reading of sentences like (2-a) exists, in addition to the weak one. As mentioned, their results have been replicated with different tasks (Clifton Jr & Dube 2010, Potts et al. 2015, Gotzner & Benz 2015).

The results of these studies can be summarised as follows: there is evidence for the strong inference of *every* embedding *some* (Chemla & Spector 2011 and others), there is no evidence for a difference between the latter case and the case of *no* embedding *all* (Chemla 2009c), and the potential strong inference of *no* embedding *all*, if it exists, it is much weaker than that of *no* embedding a presuppositional trigger (Chemla 2009b). In sum, we cannot know from the results above whether the strong inference of *no* embedding *all* exists and investigating this question is the main goal of the present study, which we will outline in the next sections. In particular, in our study, we develop an extension of the paradigm by Chemla (2009c) by using all four conditions above, along the lines of Chemla & Spector (2011), in order to compare the case of *every* to that of *no*. The novel contribution of the two experiments presented here is that they compare directly the two quantifiers case and compare each to a baseline control, both incompatible with the assertion (Experiment 1) and compatible with it (Experiment 2). This set up allows us to fully test the existence of the strong inference in (2-b).

3 Experiment 1

3.1 Methods

3.1.1 Participants

We recruited 60 Participants with U.S. IP addresses via Amazon’s Mechanical Turk and screened them for native language. 59 native English participants (27 male, 31 female, mean age 35.6) were included in the final analysis. Participants were equally spread across four survey versions with different orders. Participants received 10 cents for participation in the study.

3.1.2 Materials and Procedure

The experimental survey started with questions concerning age, gender and educational level. In the main part of the experiment, participants saw pairs of sentences and statements, that is, the candidate inferences. Their task was to indicate to which extent they would infer a statement from a given sentence. We used the same instructions as Chemla (2009c) with minor modifications. Participants were told to

base their judgement on their own intuition and were given two example sentences in which a statement clearly followed and one in which it did not clearly follow.

In the main part of the experiment, participants saw sentences like (4-a) and (5-a) paired with their corresponding strong and weak candidate inferences (see (4-b), (4-c) and (5-b), (5-c)) as well as true and false control conditions. Hence, the experiment used a 2 (quantifier: *every* vs. *no*) x 4 design (inference condition: literal, false, weak, strong). We asked participants to what extent they would infer a given inference on a scale from 0 to 100%, with 0% representing that a statement did not follow from a sentence and 100% that it definitely followed. For convenience, we present the corresponding statement of the same sentence for each condition (see Appendix C for all experimental materials).

- | | | | |
|-----|----|---|--------|
| (4) | a. | Every student read some of the books | |
| | b. | \neg <i>No student read all of the books</i> | STRONG |
| | c. | \neg <i>Not every student read all of the books</i> | WEAK |
| | d. | \neg <i>At least one student read some of the books</i> | TRUE |
| | e. | \neg <i>Not every student read some of the books</i> | FALSE |
| (5) | a. | No student read all of the books | |
| | b. | \neg <i>Every student read some of the books</i> | STRONG |
| | c. | \neg <i>Some student read some of the books</i> | WEAK |
| | d. | \neg <i>At least one student didn't read all of the books</i> | TRUE |
| | e. | \neg <i>Some student read all of the books</i> | FALSE |

Below, we present an example trial that participants saw with sentence (4-a) in the true control condition.

"Every student read some of the books"
suggests that
At least one of the students read some of the books
 ___ % YES
 0 % = definitely not, 100 % = definitely yes

In the true condition, participants were expected to accept the given statement (i.e., judgements close to 100 %) whereas in the false condition they should clearly reject the statement (i.e., judgements close to 0 %). Note that our true condition is an entailment which should be judged as true no matter which reading participants adopt. In the critical weak and strong conditions, participants judged the candidate inferences. Each participant judged one item per condition and was presented with all eight conditions. We created four survey versions with pseudo-randomized orders of the items and a given participant only took one of the survey versions.

3.2 Results

Figure 1 shows the mean % of yes responses across quantifier and inference conditions. We computed a series of mixed models with the factors quantifier (*no* vs. *every*), inference condition, interactions between the two factors as well as a random factor for participants. We set the strong condition of the quantifier *every* as the reference level to test whether the strong condition differed from the weak as well as the false condition.

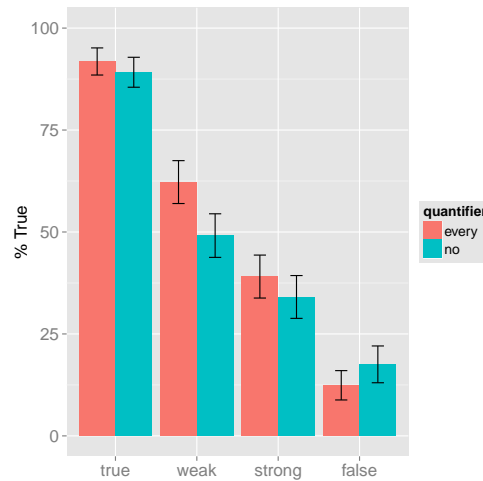


Figure 1 Mean % true responses for quantifier by inference condition (Experiment 1). Error bars represent SEM.

The model showed that the strong condition differed significantly from the true ($SE = 5.95, t = 8.86, p < .0001$), weak ($SE = 5.95, t = 3.89, p < .0001$) as well as the false conditions ($SE = 5.95, t = -4.48, p < .0001$). The difference between the two quantifier conditions was not significant at the baseline level of the strong inference condition ($p = .4$).¹⁰ Crucially, there was no significant interaction between the quantifier and inference conditions (strong vs. weak across quantifiers: $p > .34$ and strong vs. false: $p > .23$; the same pattern was found when we excluded eight participants who gave inconsistent responses in the true and false conditions). Therefore, while the strong condition differed from the weak and false ones in both quantifiers,

¹⁰ A post-hoc analysis revealed that participants endorsed the weak inference significantly more in the condition *every* than with *no* ($p < .05$). Crucially, however, there was no significant interaction between the quantifier and inference conditions. That is, participants reliably distinguished the weak and strong inference conditions with *every* as well as with *no*.

there was no evidence for an asymmetry in these differences.¹¹ Details of the mixed model are presented in Table 1. A post hoc test which tested the differences across inference conditions in each quantifier individually further confirmed that the difference between the strong and false condition was significant, both for *every* and *no* (all p -values $<.01$).

	Estimate	Standard Error	t -value	p -value
(Intercept)	39.068	4.601	8.491	0.000
No	-5.000	5.953	-0.840	0.401455
False	-26.661	5.953	-4.479	0.000
True	52.746	5.953	8.860	0.000
Weak	23.169	5.953	3.892	0.000116
No:False	10.136	8.419	1.204	0.229324
No:True	2.356	8.419	0.280	0.779743
No:Weak	-8.119	8.419	-0.964	0.335450

Table 1 Results of mixed effects model including estimates, standard errors, t -value and p -values ($n = 472$, log-likelihood = -2334.1)

3.3 Discussion

We directly compared sentences like (1-a) and (2-a) to investigate the status of their potential strong inferences in (1-b) and (2-b), respectively. Our results replicate the result by Chemla & Spector (2011) finding that the strong condition was differing from the weak condition in the *every* case. Moreover, we showed that participants also differentiate their judgements from the false no-inference control condition. The crucial finding of our study is the parallel differences across inference conditions in the cases with the quantifier *no*. Note that we did not simply observe a null effect, rather the difference between the strong and false condition and the strong and weak condition was present for sentences like (4-a) as well as (5-a). Before we move to discussing the implications of these findings, we address a potential concern with the current study, which is the main motivation for conducting Experiment 2.

In particular, the worry has to do with the nature of the baseline control. The main interest of our study was to show that the strong reading exists both for (1-a) and (2-a). Or at least, that there is as much evidence for the strong inference of (2-b) as there is for that of (1-b). Experiment 1 showed that there was a difference between the weak and strong conditions and the strong condition and a false-no

¹¹ Including the interaction term did not improve the model fit compared to the model with simple main effects for quantifier and inference condition ($\chi(3) = 4.83$, $p = .18$).

inference condition. Crucially, these differences were observed both for the *every*-sentences like (1-a) as well as the *no*-sentences like (2-a). These findings suggest that the strong readings exist in both cases. However, one may argue that the comparison between the false and the strong conditions is not a comparison that fully allows us to test whether the strong inferences exist. This is because in the false condition participants are asked to judge a potential inference that is incompatible with the assertion, whereas in the strong condition the potential inference is of course compatible with it. Therefore, if the strong inference is not an inference but it is just compatible with the assertion, participants could have just based their judgements on the plausibility of the different statements rather than actually deriving an inference. This would have still created a difference with respect to the false condition, just because the latter is instead not compatible with the assertion. That is, even in the case in which both are not inferences of the sentence in question, it could be that participants' scores reflect the fact that one is contradicted by the assertion, while the other is not.¹² We therefore ran a second experiment in which we replaced the false condition with a new control condition which was compatible with the utterance (but was not an inference on any account).

4 Experiment 2

The goal of Experiment 2 was to further investigate the strong and weak inferences of (1-a) and (2-a). Given the worry with the baseline outlined above, we aimed at finding the strongest possible control condition to test whether participants are actually computing the strong inference. If the strong condition differs from a condition which is only compatible with the assertion of the sentence (and equally complex and equally relevant), then we have evidence that such a strong condition involves an actual inference and is not just a statement compatible with the assertion. This was ensured by having the new control condition being the negation of the potential strong inference. Given the standard assumption that relevance is closed under negation, relevance cannot distinguish between the two alternatives: if one is relevant, the other has to be relevant as well (Chierchia et al. 2012, Fox 2007, Fox & Katzir 2011 among many others). Moreover, both the new control as well as the strong condition were compatible with the assertion and both were of equal complexity. The rationale is that if any of the two is a real inference of the utterance (not just compatible with it) we would observe a difference between the two.

We made a few other changes to our experimental design. First, we increased the number of items to further investigate the interaction between quantifier and inference condition. Second, in order to more directly compare our study to that of

¹² We come back to this issue in section 4.3.2 below.

Chemla (2009c), we adapted more closely his experimental materials, in particular adding a context sentence and using just “one” in the restrictor of the quantifiers (e.g. *every one*).

4.1 Methods

4.1.1 Participants

We recruited another set of 60 Participants with U.S. IP addresses via Amazon’s Mechanical Turk and screened them for native language. 59 native English participants (33 male, 26 female, mean age 33.9) were included in the final analysis. Participants received 40 cents for participation in the study.

4.1.2 Materials and Procedure

Experiment 2 used the same conditions as Experiment 1, except for the false control condition. In particular, we replaced the latter condition with a condition which was compatible with the assertion of the utterance. In (6-a) and (7-a), we illustrate an example of the compatible condition for *every* and *no* in conjunction with the strong condition. As mentioned, both conditions are compatible with the assertion, they are of comparable complexity, and one is the negation of the other. Therefore, a difference in ratings between the STRONG and the COMPATIBLE condition would be evidence that the former is an actual inference of the sentence and not just compatible with it. Moreover, any interaction between STRONG/COMPATIBLE and *no/every* would tell us whether there is any asymmetry between the two quantifiers.

- | | | | |
|-----|----|---|------------|
| (6) | a. | Every student read some of the books | |
| | b. | \rightsquigarrow No student read all of the books | STRONG |
| | c. | \rightsquigarrow Some student read all of the books | COMPATIBLE |
| (7) | a. | No student read all of the books | |
| | b. | \rightsquigarrow Every student read some of the books | STRONG |
| | c. | \rightsquigarrow Some student did not read any of the books | COMPATIBLE |

In this experiment, we created a set of four items per condition which closely resembled the items used in Chemla (2009c). In particular, we added a context sentence which introduced a noun phrase (e.g. *the students*). In the critical sentence, either *every one* or *no one* was used rather than a repetition of the noun phrase. An example stimulus is given in (8-a) (a full list of the materials is available in Appendix C).

- (8) a. Context: The students had several historical dates to remember.

The scalar inferences of negative quantifiers

- b. Every one remembered some of the dates.
- c. \neg Not every one remembered all of the dates WEAK

All conditions were presented within subject but between items. Participants saw 32 items in total. We randomized the order of items for each participants with a function build into the HTML template. The instructions were the same as in Experiment 1.

4.2 Results

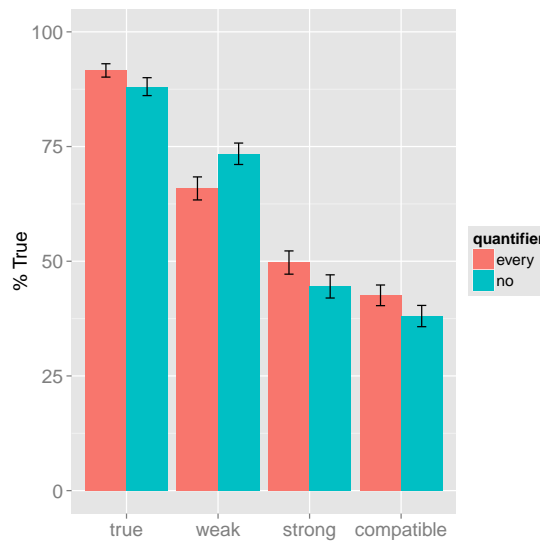


Figure 2 Mean % true responses for quantifier by inference condition (Experiment 2). Error bars represent SEM.

Figure 2 shows the mean % of yes responses across quantifier and inference conditions. We computed a series of mixed models with the factors quantifier (*no* vs. *every*), inference condition, interactions between the two factors as well as a random factor for participants and trial number. We set again the strong condition of the quantifier *every* as the reference level to test whether the strong condition differed from the weak as well as the compatible condition.

The model showed that the strong condition differed significantly from the true ($SE = 5.95, t = 14.04, p < .0001$), weak ($SE = 5.42, t = 3.89, p < .0001$) as well as the compatible conditions ($SE = 2.97, t = -2.37, p < .05$). The difference between the two quantifier conditions was marginal at the baseline level of the strong inference condition ($p = .08$). Crucially, there was no significant interaction between the

quantifier and inference conditions, comparing the strong and compatible condition (>.88). In this experiment, there was an interaction between the weak and strong condition ($SE = 4.22$, $t = 3.01$, $p < .01$), reflecting that participants accepted the strong condition in the negative quantifier to a greater extent.¹³ Details of the mixed model are presented in Table 1. A post hoc test further confirmed that all comparisons across inference conditions were significant in both quantifiers (all p -values $< .05$).

	Estimate	Standard Error	t -value	p -value
(Intercept)	49.6961	2.6421	18.809	0.0001
No	-5.1457	2.986	-1.723	0.08502
Compatible	-7.0862	2.986	-2.373	0.01775
True	41.9351	2.986	14.044	0.0001
Weak	16.2075	2.9926	5.416	0.0001
No:Compatible	0.6344	4.2182	0.15	0.88048
No:True	0.3021	4.4187	0.068	0.9455
No:Weak	12.7074	4.2205	3.011	0.00264

Table 2 Results of mixed effects model including estimates, standard errors, t -values and p -values ($n=1809$, REML = 17764.1).

4.3 Discussion

4.3.1 Summary of results and implications

The main interest of our study was to investigate whether the differences across inference conditions are present in both quantifier conditions, thereby providing evidence that the weak and strong inferences exist in parallel. Our second experiment showed that the strong condition received higher ratings than our new control condition which was just compatible with the assertion. Recall that the latter is the negation of the former, which ensures that relevance was kept constant across the two. Therefore, when we compare the strong and the control conditions, we are looking at two potential inferences both compatible with the assertion, equally complex and not differing in relevance. Hence, the fact that the former received higher ratings strongly suggest that it is because this inference is an actual inference and not just

¹³ To further process this interaction, we ran a model with the weak condition (quantifier *every*) as reference level. The model showed that participants endorsed the weak inference more in the condition with *no* than with *every* ($p < .05$), contrary to what we found in Experiment 1. We will discuss these differences below. Again, the crucial part is that the differences between the weak and strong conditions were observed for both quantifiers.

compatible with the assertion. In other words, our results provide evidence that (6-a) and (7-a) trigger the strong inferences in (6-b) and (7-b), respectively.

All in all, our results show (i) evidence for both strong inferences in (6-b) and (7-b) and (ii) no significant difference between them. In other words, we replicate Chemla & Spector's (2011) results for the case of *every* embedding *some* and, we find parallel evidence for the case of *no* embedding *all*, with the strong condition being equally accepted in both quantifiers (similar to what Chemla 2009c found).¹⁴ The main contribution of this study is that, unlike Chemla (2009c), we provide two different points of comparison for the strong condition, a false and a compatible condition, which together constitute evidence that the strong inference is indeed an inference of the sentence in question, both in the case of *every* and in that of *no*.

To further compare the differences across quantifiers, we computed the correlations of mean subject ratings across conditions (averaging over different items). This analysis showed that participants' responses in the condition with *every* and *no* were positively correlated (Pearson's correlation: 0.71, $p < .0001$ (weak condition) and .54, $p < .0001$ (strong condition). That is, participants were more likely to derive the respective inference in one quantifier condition if they had done so in the other. This further supports our main idea that the strong and weak inferences from both quantifiers were derived in a similar fashion. In addition, there was also a correlation between participant's judgements of the weak and strong inference conditions (Pearson's correlation: .35, $p < .01$ (*every*) and .34; $p < .01$ (*no*)). In other words, participants were more likely to draw a strong inference if they also derived the corresponding weak one and vice versa. Interestingly though, this relationship held to a lower extent than the one between the two quantifier conditions. For us, the most relevant aspect is that participants' derived the weak and strong inferences in a similar way with *every* and *no*.

Finally, we focused here on the strong inferences of the two quantifiers. We notice, however, that the comparison between the weak conditions showed mixed results; while in the first experiment the weak inference of *every* received a slightly higher rating than that of *no*, the second experiment showed the opposite pattern (while in the Chemla (2009c) study there was no difference between the two quantifier conditions). On the other hand, as discussed above, the strong inferences did not show the same fluctuation: the ratings of the strong inferences of *no* and *every* did not differ significantly in either experiment. Further, the differences between the strong and weak conditions were observed for each individual quantifier condition

¹⁴ Note that there was a marginal difference at the baseline level with the strong condition in Experiment 2, which was not significant in Experiment 1. Therefore, from this result we cannot conclude that there is any difference across quantifiers. If the trend of Experiment 2 would turn out to be reflecting a real difference, this would indicate, in combination with our results, that while both strong inferences exist, the strong inference of *every* embedding *some* is stronger than that of *no* embedding *all*.

in both experiments.

In the next subsection, we move to addressing various methodological issues that the type of experiments above raise. After that, we turn to discussing how the results above relate to theories of alternatives and theories of scalar implicatures.

4.3.2 Methodological issues

As extensively discussed in the literature, a natural question in relation to experiments like the above is what the differences across inference conditions reflect and how participants engage in the task. The linking assumption adopted in the original study by Chemla & Spector (2011) was that rates of endorsement of a sentence S given a picture are proportional to the number of readings of S that are true. This linking assumption can be naturally adapted to our inferential task as follows. Consider a sentence S and two inferences I_1 and I_2 : if the number of readings of S giving rise to I_1 is a subset of those giving rise to I_2 , then the degree to which I_2 will be judged to follow from S will be higher than that of I_1 .

In our results, like Chemla & Spector (2011), we indeed found a graded pattern of responses across the weak and strong conditions (though the direction of the effect was reversed, given the nature of the task, as was the case in Chemla 2009c). This suggests that participants perceived the differences across the different readings of the sentence and are, as expected, more reluctant to endorse a strong inference compared to a weaker one. And importantly, as discussed, participants judged the strong inference higher than a false statement as well as a statement that was not an inference but compatible with the assertion of the sentence.

An alternative explanation is assuming that the differences across conditions reflect deviance from the prototypical situation, as argued by Geurts & van Tiel (2013) and van Tiel (2014). As van Tiel (2014) showed typicality and contrast may be a crucial factor driving Chemla & Spector's (2011) effect, especially since the pictures used in this study made certain contrasts more salient than others. It is harder to see, however, how a typicality explanation would apply to the inferential task used here: in the inferential task, participants are directly presented with the candidate inferences and judge the degree to which it follows from the statement. It is unclear how typicality would lead to endorse one inference more than another.

While the hypothesis above can account for all the differences across conditions, it is true, on the other hand, that if we compare rates of endorsements across our two experiments, the number of possible readings alone cannot account for the difference between the FALSE and the COMPATIBLE conditions across the two experiments (to the extent that we take as indicative a comparison of percentages across different experiments).¹⁵ This is because, in both cases, the potential inference is not an

¹⁵ Thanks to an anonymous reviewer for discussion on this point about the difference between the

inference of the sentence, and therefore the hypothesis above would expect the same (very low) ratings of response. And yet, we found higher rates of endorsement of the inference in the COMPATIBLE condition in Experiment 2 compared to the one in the FALSE condition in Experiment 1. However, as we argued above, it is plausible that in addition to the number of possible readings, participants will also take into account whether a statement is contradicted entirely by the truth conditions of the assertion or whether it is compatible with the assertion, once they consider the latter as true. This, we think, is what is at the basis of the difference we observe across Experiment 1 and 2 for the FALSE and COMPATIBLE conditions. And this is also where, we think, our experiment improves on Chemla's (2009c), which did not have a baseline altogether and thereby could not distinguish between a judgment based on compatibility with the assertion versus a judgment based on an inference intuition.

Another methodological issue worth discussing relates to the use of graded judgements. Following Chemla & Spector (2011) we made use of continuous judgements, mainly because, in combination with the hypothesis above, we expected this to allow us to see a difference between the strong and weak conditions and the strong and the false or compatible ones. On the other hand, a issue raised by the nature of this type of judgements is what their exact interpretation should be. As mentioned above, Chemla & Spector (2011) discuss a natural interpretation of participants scores as being proportional to the number of possible true readings the sentence can have. And as discussed above, this for us has to be complemented with the assumption that a potential inference contradicted by the assertion is scored lower than a potential inference that is compatible with the assertion. Still, the question remains as to what exactly underlies a judgement of a relative response to a potential inference of a sentence (i.e. this is 60% an inference of the sentence), as opposed to a yes/no binary choice (i.e. this is/is not an inference of the sentence). Notice in this respect, however, that Potts et al. (2015: section 6 and Appendix C) conducted their experiments both employing a binary response task and a Likert scale and found qualitatively identical results, replicating the results found by Chemla & Spector (2011). This suggests that the effects investigated in these studies are independent from the choice between continuous and binary judgements.

Finally, another concern that has been raised in relation to inferential tasks is that they may produce an inflated number of implicatures (Geurts & Poussoulous 2009). However, while this can inflate the effect of implicatures overall, it is unclear that it can distinguish among the different conditions involving an inference. And the fact that participants differentiated all experimental conditions provides evidence for the different inferences discussed above. And more crucial for us, the evidence we found for the strong inference in the cases with *no* and *every* were entirely parallel.¹⁶

FALSE and COMPATIBLE conditions and the point below about binary versus continuous judgements.

¹⁶ Notice that the question whether the strong reading is computed by default is orthogonal to the the

5 General discussion

The results of Experiment 1 and 2 together strongly suggest that theories of scalar implicatures and alternatives need to derive the inference in (2-b) from sentences like (2-a). In this section, we discuss how the inference in (2-b) can be derived and how this interacts with issues related to under- and over-generation problems pointed out for theories of how the alternatives of multiple scalar terms sentences combine and the question as to whether embedded scalar implicatures exist.

5.1 Deriving the inferences allowing all possible alternatives

A standard way to think about the alternatives of simple sentences like *Some student came* is to assume that terms like *some* are associated with others in a scale or a set (Horn 1972, Gazdar 1979, Sauerland 2004b).¹⁷ In particular, the idea is that *some* and *all/every* are associated to each other and when we have a sentence containing one, we can construct a sentence containing the other as its alternatives by replacing the two scalar terms.¹⁸ The question is what happens when we have sentences like (1-a) or (2-a) above in which we have more than one scalar term. The standard response to this question is assuming that the alternatives of a sentence with two or more scalar terms like are simply the combination of all alternatives associated with each scalar term (Horn 1972, Sauerland 2004b, Magri 2010, Chierchia 2004, Klinedinst 2007). We can formulate this as in (9).

- (9) **Alternatives:** The set $\mathcal{Alt}(\phi)$ contains all and only those ψ 's that can be obtained from ϕ by replacing one or more scalar items in ϕ with their scale-mates.

Given (9), we can see now that the alternatives that we can generate from (1-a) by replacing each scalar term with each scale-mate are those in (10).

$$(10) \quad \mathcal{Alt}(1-a) = \left\{ \begin{array}{l} \text{Every student read some, Every student read all} \\ \text{Some student read some, Some student read all} \end{array} \right\}$$

In addition to a mechanism for generating alternatives like (9), it is generally assumed that the context can restrict the set of generated alternatives to a subset of *relevant*

question whether a certain reading exists (see (Gotzner & Benz 2015) for a more detailed discussion and evidence that the strong reading is the preferred interpretation in certain contexts).

¹⁷ See Katzir 2007 and Fox & Katzir 2011 for a theory of alternatives that doesn't rely on a notion of 'scalar term.' For our purposes, as far as we can see, nothing changes with respect to the problems discussed below if we adopt a theory of alternatives like Katzir 2007, Fox & Katzir 2011.

¹⁸ We will use in the text *every* and *all* in the text below interchangeably, disregarding subtle differences between them, which are irrelevant for our purposes.

ones. We will not try to define relevance precisely, but we will assume that the set of relevant alternatives $\mathcal{Alt}_R(\phi)$ is $\mathcal{Alt}(\phi) \cap \mathcal{R}$, where \mathcal{R} is the set of propositions that are relevant in the context. (10) is then combined with a theory of how scalar implicatures are generated on the basis of alternatives. For our purposes, we can assume a formulation like (11) (Sauerland 2004b, Fox 2007 a.o.).¹⁹

- (11) **Scalar implicature derivation:** the enriched meaning of a sentence ϕ , call it $\mathcal{S}(\phi)$, arises by conjoining the literal meaning and the negation of any alternative q in $\mathcal{Alt}_R(\phi)$ such that:
- a. ϕ doesn't entail q and
 - b. $\neg q \wedge \phi$ does not entail any other $p \in \mathcal{Alt}_R(\phi)$

Given (9) and (11) and the assumption about relevance, we can now show how the inferences above can be generated. First, as we saw the alternatives of (1-a) generated by (9) are those in (10) above. If we assume that the set of relevant alternatives in the context only include (12) and we apply our scalar implicature algorithm in (11), we end up negating that every student read all of the books, thereby obtaining the weak inference in (13).

$$(12) \quad \mathcal{Alt}_{\mathcal{R}}(1-a) = \left\{ \begin{array}{l} \text{Every student read some of the books} \\ \text{Every student read all of the books} \end{array} \right\}$$

$$(13) \quad \mathcal{S}(1-a) = \text{every student read some} \wedge \neg(\text{every student read all})$$

If the set of relevant alternatives is (10), instead, we obtain the stronger inference in (14), negating the non-weaker alternative that some student read all of the books.

$$(14) \quad \mathcal{S}(1-a) = \text{every student read some} \wedge \neg(\text{some student read all}) = \text{every student read some} \wedge \text{no student read all.}$$

What is important for us is that in a completely parallel fashion, using the same ingredients above, we can obtain the inferences of (2-a). First, we generate the set of alternatives in (15).²⁰

¹⁹ Notice that (11) allows for the negation of alternatives that are logically independent from the assertion, in addition to the ones that are strictly stronger than the latter. For arguments in favour of this step, see Spector 2007, Chierchia et al. 2012, Romoli 2012 among others.

²⁰ See Romoli 2012 for arguments for why *no* and *not every* are alternatives to each other. In brief, the arguments are that generally *no* is assumed to be decomposed as *not some*. If this is so, given the assumption about alternatives above, we can create *not every* alternatives, by replacing *some* with *every* under negation. Moreover, a sentence like (i-a) has the inference in (i-b), which is standardly assumed to be derived by negating the alternative in (i-c). This shows that *no* has to be a scale-mate with *not every* and therefore if the scale-mate relation is symmetric as generally assumed also *not every* is a scale-mate of *no*.

$$(15) \quad \mathcal{Alt}(2-a) = \left\{ \begin{array}{l} \text{No student read all of the books} \\ \text{No student read some of the books} \\ \text{Not every student read all of the books} \\ \text{Not every student read some of the books} \end{array} \right\}$$

If we assume that the set of relevant alternatives in the context are only (16) and we apply our scalar implicature algorithm, we end up negating that no student read some of the books thereby obtaining the weak inference in (17).

$$(16) \quad \mathcal{Alt}_{\mathcal{R}}(2-a) = \{ \text{No student read all, No student read some} \}$$

$$(17) \quad \mathcal{S}(1-a) = \text{no student read all} \wedge \neg(\text{no student read some}) = \\ \text{no student read all of the books} \wedge \text{some student read some of the books.}$$

If the set of relevant alternatives is, instead, identical to (15), we obtain the inference in (18), negating the alternative that not every student read some of the books.

$$(18) \quad \mathcal{S}(1-a) = \text{no student read all} \wedge \neg(\text{not every student read some}) = \\ \text{no student read all} \wedge \text{every student read some.}$$

In sum, given standard assumptions about how alternatives are constructed and combine and how scalar implicatures are generated, we obtain the inferences of sentences like (1-a) and (2-a) in parallel ways.²¹

We turn now to discuss how the debate between local vs global scalar implicatures touches on the issues discussed above. Notice that the answer to the question as to whether there are genuine embedded scalar implicatures has far-reaching implications for the theory of scalar implicatures and the semantics-pragmatics interface. We will not try to answer this question here, but will only show how the assumption that there are embedded scalar implicatures interact with assumptions about the theory of alternatives for the cases under discussion here.

-
- (i) a. Not every student came.
 b. Some student came.
 c. No student came.

²¹ One open question for this approach, of course, is what constraints the selection of one set of alternatives versus the other, with the corresponding different inferences, and more would have to be said than the simplified notion of relevance given here. In any case, as we will show below, this approach is inadequate to deal with other cases with multiple scalar terms.

5.2 Global versus local implicatures

In the section above, the scalar implicature algorithm was always only applying at the global level of the sentences in question. As mention, there is, however, a debate as to whether such mechanism can also apply in embedded positions. This is relevant here in particular because the strong inference of (1-a) could also be obtained if one allows the scalar implicature algorithm to apply locally as in (19) with respect to the alternatives of that part of the sentence in (20).

(19) Every student₁ $\mathcal{S}[t_1$ read some of the books]

(20) $\mathcal{Alt} = \{ t_1$ read some of the books, t_1 read all of the books $\}$

The result of enriching (1-a) as in (19) is the strong inference in (21).

(21) Every student read some of the books and not all of the books =
Every letter student read some of the books and no student read all.

Notice that the same local option is not available for (2-a), as local enrichment as in (22) is vacuous: all alternatives in (23) are weaker, so no inference is derived.²²

(22) No student₁ $\mathcal{S}[t_1$ read all of the books]

In sum, given a standard approach to the alternatives of multiple scalar terms, we obtain the weak and strong inferences of (1-a) and (2-a) in a parallel fashion. In addition, (1-a), unlike (2-a), might also allow another way of drawing the strong inference, by locally applying the scalar implicature algorithm. What is important for us here is that the inference of (2-a) cannot be derived locally. In the next subsection, we go back to the global derivations of these inferences and show that the simple standard idea above over- and under-generate in other cases, and therefore need to be constrained somehow. We then discuss how constraining the theory of alternatives interact with the derivation of the inference of (2-a).

5.3 The need for constraining the combination of alternatives

We saw above that the standard way of thinking about how the alternatives of sentences with multiple scalar terms are derived can account for the inferences above. This theory, however, has been shown to be problematic, in that (9), coupled

²² More precisely, one could obtain the inference locally if one were to decompose *no* as *every ... not* and apply our algorithm above *not*, but this decomposition is ad hoc and quite problematic in various respects; see Chemla 2009c for discussion.

(i) Every student₁ $\mathcal{S}[not[t_1$ read some of the books]]

with the a mechanism for deriving scalar implicatures like (11), gives rise to both an under- and an over-generation problem (Fox 2007, Magri 2011, Romoli 2012, Chemla 2010, Trinh & Haida 2015 among others). In particular, it has been shown that (9) under-generates with sentences like (23-a), in that it doesn't predict the intuitively correct inferences in (23-b) and (23-c), while it over-generates with sentences like (24-a) in that it predicts the inference in (24-b), which, instead, is not intuitively a possible inference for (24-a).

- (23) a. Every student took Syntax or Semantics.
 b. \rightsquigarrow *Some student took Syntax*
 c. \rightsquigarrow *Some student took Semantics*
- (24) a. Some student read all of the books.
 b. $\not\rightarrow$ *Some student didn't read any of the books*

To see the problem with (23-a), consider the alternatives that (9) allows us to construct, by replacing each of the two scalar terms in the sentence and take the crossproduct of all replacements, (25).²³

- (25) $\left\{ \begin{array}{l} \text{Every student took Syntax or Semantics, Every student took Syntax,} \\ \text{Every student took Semantics, Every student took Syntax and Semantics,} \\ \text{Some student took Syntax or Semantics, Some student took Syntax,} \\ \text{Some student took Semantics, Some student took Syntax and Semantics} \end{array} \right\}$

The problem here is that if the relevant alternatives include all alternatives in (25), we cannot obtain the inferences in (23-b) and (23-c) by negating the alternatives that *every student took Syntax* and *every student took Semantics*. This is because (25) contains alternatives that are equivalent to the inferences that we want to obtain, so we will never be able to obtain them given (11).

To see the problem with (24-a) consider the alternatives that we can construct for it in (26) and assume they are all relevant: applying the scalar implicature algorithm we would incorrectly obtain the inference in (27), by negating the alternative that all students read some of the books.

- (26) $\mathcal{Alt}(24-a) = \left\{ \begin{array}{l} \text{Some student read all of the books} \\ \text{Some student read some of the books} \\ \text{All student read all of the books} \\ \text{All student read some of the books} \end{array} \right\}$

- (27) Some student read all of the books \wedge \neg (all student read some) =

²³ We are assuming here, following Sauerland (2004b) and much subsequent work that a disjunction has its disjuncts as alternatives.

The scalar inferences of negative quantifiers

Some student read all of the books and some student didn't read any.

5.4 Fox's proposal

5.4.1 The basic proposal

Based in particular on the data about the under-generation problem, Fox (2007) (see also Magri 2011, Romoli 2012 and Chemla 2010) has proposed the more constrained hypothesis of how the alternatives of different scalar terms combine in (28). (28) requires that alternatives can only be obtained by replacing one scalar term at a time and only if the resulting alternative is not entailed by the previous step.

- (28) **Fox's hypothesis:** The set $\text{Alt}(\phi)$ is recursively defined as follows:
 $\phi \in \text{Alt}(\phi)$ and
 $\psi \in \text{Alt}(\phi)$ iff there is $\chi \in \text{Alt}(\phi)$ such that ψ is not weaker than χ and furthermore ψ is obtained from χ by replacing a single scalar item in ϕ with a scale-mate.

Consider now how (28) would solve both the under- and over-generation problem.

First, with (28) we cannot construct the problematic logically independent alternative in (29) for a (24-a), thereby avoiding the over-generation problem.

- (29) All of the students did some of the readings.

This is because there is no way to replace one scalar item at a time in (24-a) and construct (29), without also going from stronger to weaker alternatives.²⁴ Finally, the procedure also blocks the alternatives that create the under-generation problem of sentences like (23-a) above. Remember that the problematic alternatives were *some student took syntax* and *some student took semantics*. Consider the former: we need to make two replacements to obtain it, replacing disjunction with one of its disjuncts and replacing *every* with *some*. It is easy to see that no matter where we start from, we cannot get it with Fox's procedure.²⁵

²⁴ To illustrate, consider the options that we have in constructing alternatives from (24-a): first, we cannot replace *all* in (24-a), because we would obtain the weaker alternative (i). The only other option is replacing *some* in (24-a). In this way, we obtain the alternative in (ii), which is stronger than (24-a), therefore it is a licit alternative. (ii), however, is also stronger than (28), therefore, we know we cannot obtain (28) from (ii) (by replacing the embedded *all*).

- (i) Some of the students did some of the readings.
(ii) All of the students did all of the readings.

²⁵ This is because if we start by replacing *every* with *some* we obtain (i-a), which is weaker than (23-a).

5.4.2 A modification involving recursive implicature computation

As discussed, Fox (2007) provides a solution for both the under- and over-generation problems, but does not, as it is, predict the derivation of the inference the two experiments above show evidence for. As pointed out to us by Luka Crnic (p.c.), however, if we allow for recursive/multiple computation of implicatures, and we modify some assumptions Fox makes about the alternatives in such multiple computations, the situation changes. In particular, a possible derivation of the strong inference of *no* embedding *all*, which is compatible with Fox's (2007) approach, becomes available. The option of recursive implicature computation, however, re-opens the over-generation problem as well. In other words, if we allow ourselves to recursively compute implicatures, the inference in (2-b) can be derived in a way that is compatible with Fox's approach, but the same type of derivation makes the over-generation problem re-emerge. We illustrate both of these points in detail in Appendix A. The gist of the effect of multiple scalar implicature computation is the following. If we consider a recursive application of our implicature algorithm \mathcal{S} on a sentence like (2-a), the alternatives over which the second \mathcal{S} operates on, have already been applied the first \mathcal{S} . As a consequence, some of the previous entailment relations among alternatives will be broken. In particular, the alternative *No student read some of the books* does not entail the alternative *Not every student read some of the books*, once we apply \mathcal{S} to both. Therefore, the latter is not excluded anymore by Fox's constraint and the inference to its negation can be derived.

Similarly, however, the recursive computation of implicatures on a sentence like (24-a) has the effect of breaking entailment relations among the alternatives and in particular the entailment relation between the alternative *All of the students read all of the books* and the alternative *All of the students read some of the books*. As a consequence, the latter alternative is not blocked anymore by Fox's constraint, and the problematic inference to its negation can be derived again.

In sum, Fox's original proposal solves the under- and over-generation problems, but cannot account for the inference in (2-b) for which we provided evidence for. The modification discussed in this subsection can account for the latter but loses an account of the over-generation problem. In the next subsection, we turn to a different attempt at constraining the alternatives of multiple scalar terms by Romoli (2012).

If we start by replacing disjunction with one of its disjunct we obtain (i-b), which is stronger than (23-a). However, if we now replace *every* in (i-b) with *some* we obtain (i-c), which is weaker than (i-b), thus we cannot include it in the set of alternatives. The same applies to the other disjunct.

- (i)
 - a. Some student took syntax or semantics
 - b. Every student took syntax
 - c. Some student took syntax

5.5 Romoli's proposal

5.5.1 The basic proposal

Romoli (2012) proposes a different constraint on alternative combination.²⁶ The motivation for Romoli (2012) to develop an alternative to Fox's proposal is precisely related to the strong inference of sentences like (2-a). That is, Romoli (2012) argues that his procedure has an advantage over Fox's (2007), precisely because the latter does not allow us to derive this inference. In particular, we have shown above that assuming the traditional hypothesis in (9) about alternatives, the strong inference in (1-b), can be obtained by adding a scalar implicature obtained by negating the alternative *every student read some of the books*, in turn obtained by replacing *some* with *every* and *every* with *some*. The strong inference of (2-a) in (2-b), in parallel, can be obtained by negating the alternative *Not every student read all of the books*, in turn obtained by replacing *not every* to *no* and *some* for *every*.

However, as discussed, once we assume Fox's hypothesis above, we do not have the possibility of globally deriving (1-b) and (2-b) any more. To see this, consider again that the alternatives of (1-a) that we need for obtaining the inference in (1-b) is (30). It is clear that there is no way of constructing (30) from (1-a) by replacing one scalar term at a time and without taking a 'weakening step' at the same time.²⁷ Same reasoning applies to (2-a) and the alternative that we need for the inference in (2-b), which is (31).²⁸ Therefore, without additional assumptions, the theory of alternatives by Fox would predict no strong inferences for either (1-a) and (2-b), contrary to our results.²⁹

26 See Trinh & Haida 2015 for a similar proposal.

27 For instance, if we substitute *all* for *some*, we obtain *every student read all of the books*. This alternative is allowed by Fox's procedure and will give us the weak inference in (1-c), however from it we cannot construct (30), substituting *some* for *every*, as (30) is weaker than *every student read all of the books*.

28 Again for instance we could try to substitute *some* for *all* and obtain *no student read some of the books*. But from the latter we would not be able to obtain (31), as the latter is again weaker than *no student read some of the books*.

29 In the case of (1-a), this prediction is already challenged by the results of Chemla & Spector (2011). It is here, however, that the debate about the theory of alternatives of multiple scalar terms connects with the local vs global debate. This is because, as we discussed above, in the case of (1-a), one can assume that the strong inference is derived locally. That is, one can derive the strong inference by adding an embedded scalar implicature to (1-a). As shown above, however, there is no corresponding local way of obtaining (2-b), modulo ad hoc assumptions about the decomposition of negative quantifiers. In other words, Fox's hypothesis plus the assumption that implicatures can appear at embedded levels would predict an asymmetry between (1-a) and (2-a), in that only (1-a) should have the strong inference in (1-b), while (2-a) should not have the corresponding one in (2-b). But this prediction is disconfirmed by our experimental results.

- (30) Some student read all of the books.
 (31) Not every student read some of the books.

Given the motivation above, Romoli (2012) proposes a different constraint on alternatives formulated as the procedure in (32), which starts from the most embedded scalar term and construct the set of alternatives by going up the tree.

- (32) **Romoli’s hypothesis:** **Step 1:** Construct all possible alternatives of S by only replacing its most embedded scalar item, call this set $\mathcal{A}lt_1$
Step 2: Compute the excludable subset of $\mathcal{A}lt_1$. Call it $\mathcal{E}xcl_1$.
Step 3: Consider the set of alternatives $\mathcal{A}lt_1'$, which is $\{ \llbracket S \rrbracket \} \cup \mathcal{E}xcl_1$
Step 4: Starting from $\mathcal{A}lt_1'$, collect all possible alternatives by only replacing the *next* most embedded scalar item and obtain $\mathcal{A}lt_2$.³⁰
Step 5: Compute the excludable subset of $\mathcal{A}lt_2$. Call it $\mathcal{E}xcl_2$.
 ... Repeat until you exhaust all scalar items in S .
Final Step: the set of excludable alternatives is the last excludable set obtained with the steps above, $\mathcal{E}xcl_n$.

As Romoli shows, this procedure, like Fox’s, blocks the over-generation problem, in that it doesn’t allow us to derive the alternative in (24-a) for a sentence like (24-a). But crucially, it also allows the inference of (2-b) from (2-a), compatibly with the results of the experiments above. In other words, Romoli’s procedure both solves the over-generation problem and can account for the inference of sentences like (2-a).³¹

As Romoli (2012) himself discusses, while his proposal can account for the over-generation problem and the strong inference of (2-a) above, it has the drawback

30 The next most embedded scalar of Y in some sentence ϕ is simply the most embedded scalar item in ϕ if we ignore Y .

31 To illustrate briefly, Romoli’s constraint predicts the inference in (2-b) because it proceeds in steps, starting from the most embedded scalar term. For a sentence like (2-a), the alternatives constructed at the first stage are (i-a) and the excludable of which is *None of the students read some of the books*. This alternative is then used to construct the alternatives in (i-b) at the second step, by replacing *None* with *Not every*. From (i-b) the strong inference in (2-b) is then derived.

- (i) a. {No student read all, No student read some }
 b. { ... Not every student read all, Not every student read some }

On the other hand, the problematic alternative *All students read some of the books* for (24-a) cannot be constructed because none of the alternatives in (ii-a) obtained by replacing the most embedded scalar terms are excludable. At the subsequent step therefore only the alternatives in (ii-b) can be constructed, so that the over-generation problem is avoided.

- (ii) a. {Some student read all, Some student read some }
 b. {Some student read all, All students read all }

of not solving the under-generation problem. There has been, however, a recent proposal on how to solve the under-generation problem, which is independent from the way the alternatives of multiple scalar terms are constrained, and which could be combined with Romoli's 2012 account. We turn to briefly describe this proposal in the next subsection.

5.5.2 Adding an independent solution to the under-generation problem

Crnic et al. (2015) investigate experimentally the inferences in (23-b) and (23-c) from a sentence like (23-a) repeated in (33) and argue against the traditional approach to the derivation of these inferences above, by providing evidence of cases in which the sentence in (33) has the inferences in (33-a) and (33-b), in contexts in which, say, every student took syntax and some of them took semantics.

- (33) Every student took syntax or semantics.
a. \rightsquigarrow *Some student took syntax*
b. \rightsquigarrow *Some student took semantics*

Importantly, in such a context, the inferences in (33-a) and (33-b) cannot be obtained by negating the alternative that *every student took syntax* because the latter is actually true in the context by assumption. They nonetheless show that participants compute the inferences in (33-a) and (33-b) in such contexts.

Motivated by these data which are problematic for the tradition way of capturing these inferences, Crnic et al. (2015) propose a different derivation of (33-a) and (33-b) involving a multiple computation of alternatives and various assumptions about which alternatives need to be considered. We refer the reader for the details of Crnic et al.'s (2015) proposal.³² Here instead we turn to the recent account by Bar-Lev & Fox (2016), which also derives the inferences in (33-a) and (33-b) in a way that is compatible with Crnic et al.'s (2015) data. In addition, crucially for us, their derivation bypass the under-generation problem altogether and requires no extra specific assumptions about alternatives.

In brief, they propose that the inference should be derived by recursively computing implicatures at the global level as in (34).

- (34) $\mathcal{S}(\mathcal{S}(\text{every student took syntax or semantics}))$

As we show in Appendix B, while the first \mathcal{S} leads to no inference, the second

³² Crnic et al.'s (2015) account requires specific assumptions about the theories of alternatives in order to derive the inferences above and does not, in itself, constitute a solution to the under-generation problem when combined with Romoli's (2012) proposal. Thanks to Moshe Bar-Lev (p.c.) for discussion on this point.

application of \mathcal{S} yields the desired inferences in (33-a) and (33-b) by negating the alternatives in (35-a) and (35-b):

- (35) a. $\mathcal{S}(\text{Every student took syntax}) =$
 Every student took syntax $\wedge \neg(\text{some student took semantics})$
 b. $\mathcal{S}(\text{Every student took semantics}) =$
 Every student took semantics $\wedge \neg(\text{some student took syntax})$

The negation of (35-a) and (35-b), together with the assertion, entails that some student took syntax and some student took semantics.

In sum, if we adopt Bar-Lev & Fox's (2016) proposal, the under-generation problem is avoided independently from theories of the alternatives of multiple scalar terms and by combining that with Romoli's (2012) constraint, we can address both the under- and over-generation problem, while also deriving the strong inference of (2-a), of which we found evidence of in our experiments.

In closing, let us mention that as Crnic et al. (2015) discuss, their observation about (33) does not, at least intuitively, extend to modals, where the situation is completely parallel. In other words, while they did not test modal sentences experimentally, intuitively the sentence in (36), which gives rise to the inferences in (36-a) and (36-b), is not felicitous in the corresponding context in which (33) was felicitous, which is one in which Mary is required to take syntax and allowed to take semantics.

- (36) Mary is required to take syntax or semantics.
 a. \rightsquigarrow *Mary is allowed to take Syntax*
 b. \rightsquigarrow *Mary is allowed to take Semantics*

The problem for Crnic et al. (2015), and Bar-Lev & Fox (2016) as well, is to explain why the inferences in (36-a) and (36-b) appears to be derived in the way they suggest which would be compatible with such a context. Moreover, and most relevantly for us, if this contrast between nominal quantifiers sentences like (33) and the corresponding modal sentences in (36) entails that we can only derive the inferences in (36-a) and (36-b) in the traditional way, then the under-generation problem remains at least for sentences like (36). We leave further investigation of cases like (36) in comparison to (33) for further research.

6 Conclusion

We directly compared sentences like (1-a) and (2-a) to investigate the status of their potential inferences in (1-b) and (2-b), respectively. In our results, we replicated the result by Chemla & Spector (2011) finding evidence for the strong inference of (1-a).

Crucially, we found novel and parallel evidence for the strong inference of (2-a) in (2-b) and no difference between them. These results show that theories of scalar implicatures in combination with theories of how the alternatives of sentences with multiple scalar terms combine have to predict both of these inferences. Standard theories of alternatives are compatible with our data, but, as discussed, they incur in an over- and under-generation problem with other cases. We discussed the more constrained theories of alternatives by Fox 2007 and Romoli 2012 and we showed that the former can account for both the under- and over-generation problems, but cannot account for our data, while the latter can account for the over-generation problem, as well as our data, but doesn't address the under-generation problem. However, as we discussed, to the extent that the under-generation problem can be given an independent solution (Bar-Lev & Fox 2016), the proposal by Romoli (2012) appears at the moment the most promising way of constraining the alternatives of multiple scalar terms, in a way that account for our data and doesn't incur in the problems discussed in the literature.

References

- Abusch, Dorit. 2010. Presupposition triggering from alternatives. *Journal of Semantics* 27(1). 1–44.
- Bar-Lev, Moshe & Danny Fox. 2016. On the global calculation of embedded implicatures. In *MIT EXH workshop*, .
- Beaver, David. 1994. When variables don't vary enough. In M. Harvey & L. Santelmann (eds.), *Proceedings of SALT 4*, Cornell University: CLC Publications.
- Beaver, David. 2001. *Presupposition and Assertion in Dynamic Semantics*. Stanford University: CSLI Publications.
- Benz, Anton & Nicole Gotzner. 2014. Embedded implicatures revisited: Issues with the truth-value judgment paradigm. In Judith Degen, Michael Franke & Noah D. Goodman (eds.), *Proceedings of the formal & experimental pragmatics workshop*, 1–6. Tübingen.
- Charlow, Simon. 2009. "Strong" predicative presuppositional objects. In *Proceedings of ESSLLI 2009*, Bordeaux.
- Chemla, Emmanuel. 2009a. Presuppositions of quantified sentences: experimental data. Unpublished previous version of Chemla 2009b available at <http://www.emmanuel.chemla.free.fr>.
- Chemla, Emmanuel. 2009b. Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics* 17(4). 299–340.
- Chemla, Emmanuel. 2009c. Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics* 2(2).
- Chemla, Emmanuel. 2010. Similarity: towards a unified account of scalar impli-

- catures, free choice permission and presupposition projection. Unpublished manuscript.
- Chemla, Emmanuel & Benjamin Spector. 2011. Experimental evidence for embedded scalar implicatures. *Journal of Semantics* 28(3). 359–400.
- Chemla, Emmanuel & Benjamin Spector. 2015. Distinguishing typicality and ambiguities, the case of scalar implicatures. Unpublished manuscript ENS.
- Chierchia, Gennaro. 1995. *Dynamics of meaning*. Chicago: University of Chicago Press.
- Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Adriana Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, vol. 3, 39–103. Oxford: Oxford University Press.
- Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Claudia Maienborn, Klaus von Stechow & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning volume 3*, Berlin: Mouton de Gruyter.
- Clifton Jr, Charles & Chad Dube. 2010. Embedded implicatures observed: A comment on. *Semantics and pragmatics* 3(7). 1.
- Crnjic, Luka, Emmanuel Chemla & Danny Fox. 2015. Scalar implicatures of embedded disjunction. *Natural Language Semantics* 23(4). 271–305.
- Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, 71–120. Palgrave.
- Fox, Danny. 2012. Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening. MS the Hebrew University of Jerusalem.
- Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107.
- Gazdar, Gerald. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.
- Geurts, Bart & Nausicaa Pouscoulous. 2009. Embedded implicatures?!? *Semantics and pragmatics* 2. 4–1.
- Geurts, Bart & Bob van Tiel. 2013. Embedded scalars. *Semantics and Pragmatics* 6(9). 1–37.
- Gotzner, Nicole & Anton Benz. 2015. The best response paradigm and a comparison of different models of implicatures of complex sentences. Ms.
- Heim, Irene. 1982. *The semantics of definite and indefinite Noun Phrases*: University of Massachusetts, Amherst dissertation.
- Horn, Lawrence. 1972. *On the semantic properties of logical operators in English*:

- UCLA dissertation.
- Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistic and Philosophy* 30(6). 669–690.
- Klinedinst, Nathan. 2007. *Plurality and possibility*: UCLA dissertation.
- Magri, Giorgio. 2010. *A theory of individual-level predicates based on blind mandatory scalar implicatures*: Massachusetts Institute of Technology dissertation.
- Magri, Giorgio. 2011. Another argument for embedded scalar implicatures based on oddness in DE environments. In David Nan Li & Lutz (ed.), *Semantics and Linguistic Theory (SALT) 20*, Vancouver, British Columbia.
- Meyer, Marie-Christine & Uli Sauerland. 2009. A pragmatic constraint on ambiguity detection. *Natural Language and Linguistic Theory* 27(1). 139–150.
- Potts, Christopher, Daniel Lassiter, Roger Levy & Michael C. Frank. 2015. Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. <http://dx.doi.org/10.1093/jos/ffv012>. Ms., Stanford and UCSD.
- Romoli, Jacopo. 2012. *Soft but strong: Neg-raising, soft triggers, and exhaustification*: Harvard University dissertation.
- Romoli, Jacopo. 2014. The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of semantics* 0.
- Rothschild, Daniel. 2011. Explaining presupposition projection with dynamic semantics. *Semantics and Pragmatics* 4(3). 1–43. <http://dx.doi.org/10.3765/sp.4.3>.
- Russell, Benjamin. 2012. *Probabilistic reasoning and the computation of scalar implicatures*: Brown University dissertation.
- Sauerland, Uli. 2004a. On embedded implicatures. *Journal of Cognitive Science* 5. 107–137.
- Sauerland, Uli. 2004b. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27(3). 367–391.
- Sauerland, Uli. 2010. Embedded implicatures and experimental constraints: A reply to geurts & pouscoulous and chemla. *Semantics and Pragmatics* 3. 2–1.
- Schlenker, Philippe. 2009. Local contexts. *Semantics and Pragmatics* 2(3). 1–78. <http://dx.doi.org/10.3765/sp.2.3>.
- Simons, Mandy. 2001. On the conversational basis of some presuppositions. In Rachel Hastings, Brendan Jackson & Zsófia Zvolenszky (eds.), *Semantics and Linguistic Theory (SALT) 11*, 431–448.
- Spector, Benjamin. 2007. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, Palgrave.
- Sudo, Yasutada. 2012. *On the semantics of Phi features on pronouns*: MIT dissertation.
- Sudo, Yasutada, Jacopo Romoli, Danny Fox & Martin Hackl. 2011. Variation

- of presupposition projection in quantified sentences. In *Proceedings of the amsterdam colloquium 2011*, Amsterdam, The Netherlands.
- van Tiel, Bob. 2014. *Quantity matters: Implicatures, typicality and truth*. Radboud Universiteit Nijmegen dissertation.
- Trinh, Tue & Andreas Haida. 2015. Building alternatives. *Journal of Semantics* .
- Van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 137–175.
- Zehr, Jeremy, Cory Bill, Lyn Tieu, Jacopo Romoli & Florian Schwarz. 2015. Existential presupposition projection from ‘none:’ an experimental investigation. In *Pre-proceedings of the Amsterdam Colloquium 2015*, .

7 Appendix A: Crnic’s modification of Fox (2007)

In this Appendix, we show that once we assume recursive computation of implicatures, we can account for the strong inference of *no* embedding *all*, in a way compatible with Fox’s constraint on alternatives. On the other hand, once we do that, we incur in the over-generation problem again.³³ We turn to these two points in turn in the following. Before, let us define what we mean by recursive implicature computation.

In the above, we have assumed that implicatures arise through a computation which takes the form in (37).

- (37) **Scalar implicature derivation:** the enriched meaning of a sentence ϕ , call it $\mathcal{S}(\phi)$, arises by conjoining the literal meaning and the negation of any alternative q in $\mathcal{Alt}_R(\phi)$ such that:
- a. ϕ doesn’t entail q and
 - b. $\neg q \wedge \phi$ does not entail any other $p \in \mathcal{Alt}_R(\phi)$

Once we have something like \mathcal{S} , there is no technical obstacle to apply it recursively or more than once in a sentence, in the same way as there was no obstacle to apply it locally.³⁴ In other words, when we have a sentence like ϕ , we can ask what is the interpretation of ϕ if we interpret it as $\mathcal{S}(\mathcal{S}(\phi))$.

In addition, Fox (2007) assumes that the alternatives of sentences involving more than one implicature computation are defined as follows (where for each implicature computation \mathcal{S} we now indicate as subscript its set of alternatives):

³³ Thanks to Luka Crnic (p.c.) for pointing this out to us and for discussion.

³⁴ Like in the case of local applications of \mathcal{S} , recursive applications of it also raises questions about the nature of scalar implicature computation, which we will have to put aside here (see Chierchia et al. 2012, Fox 2007 and Magri 2011 among others for discussion).

The scalar inferences of negative quantifiers

(38) The alternatives A' of $\mathcal{S}_{A'}\mathcal{S}_A(\phi) = \{\mathcal{S}_A(\psi) : \psi \in A\}$

The alternatives A' of the multiple scalar implicature computation $\mathcal{S}_{A'}(\mathcal{S}_A(\phi))$ consists of each of the member of the alternatives A of $\mathcal{S}_A(\phi)$ strengthened by a first implicature computation \mathcal{S}_A . Essentially, what (38) does is constraining the second application of \mathcal{S} to only operate on alternatives over which the first application of \mathcal{S} operates on.

With this background in place, we can now go back to the main sentence above, repeated in (39), and show that we can derive the inference in (40) once we make the following two assumptions: first, we allow recursive implicatures and second, we drop the constraint on the alternatives of sentences involving multiple scalar implicature computation in (38) so that the two set of alternatives are constructed independently only following the algorithm in (28).

(39) None of the students read all of the books.

(40) All of the students read some of the books.

7.1 Deriving the strong inference

Consider the sentence in (39) and compute implicature recursively as in (41).

(41) $\mathcal{S}(\mathcal{S}(\text{None of the students read all of the books}))$.

And now allow the definition of the alternatives to just proceed independently for A' and A following the algorithm in (28).

In particular, first, consider the alternatives over which the first \mathcal{S} applies, which are in (42).

(42) $\left\{ \begin{array}{l} \text{None of the students read all of the books} \\ \text{None of the students read some of the books} \end{array} \right\}$

The result of applying \mathcal{S} over the alternatives in (42) is:

(43) $\mathcal{S}(\text{None of the students read all of the books}) =$
 None of the students read all of the books $\wedge \neg(\text{None of the students read some of the books}) =$
 None of the students read all of the books $\wedge (\text{some of the students read some of the books})$

The second application of \mathcal{S} operates on the alternatives in (44), each of which has been applied the first \mathcal{S} . Importantly, given that they are in the scope of \mathcal{S} , unlike the above, the second alternative does not entail the third and therefore it is

compatible with Fox's constraint.

$$(44) \quad \left\{ \begin{array}{l} \mathcal{S}(\text{None of the students read all of the books}) \\ \mathcal{S}(\text{None of the students read some of the books}) \\ \mathcal{S}(\text{Not all of the students read some of the books}) \end{array} \right\}$$

The alternatives in (44) are equivalent to (45) and the meaning of (41), giving rise to the strong inference we wanted, is therefore given in (46).

$$(45) \quad \left\{ \begin{array}{l} \text{None of the students read all of the books} \wedge \text{some read some} \\ \text{None of the students read some of the books} \\ \text{Not all of the students read some of the books} \wedge \text{some read some} \end{array} \right\}$$

$$(46) \quad \mathcal{S}(\mathcal{S}(\text{None of the students read all of the books})) = \\ \text{None of the students read all of the books} \wedge \text{some of the students read some} \\ \text{of the books} \wedge \neg(\text{not all of the students read some of the books} \wedge \text{some of} \\ \text{the students read some of the books}) = \\ \text{None of the students read all of the books} \wedge \text{all of the students read some of} \\ \text{the books}$$

7.2 The over-generation problem re-emerging

Recursively computing implicatures allows us to obtain the strong inference of *no* embedding *all* in a way that is compatible with Fox's approach. As we will show now, however, once we allow ourselves this option, the over-generation problem re-emerges. Consider the sentence in (47), repeated from above. As we discussed, we do not want to conclude (47-b) from (47-a) and, if we stick to one application of \mathcal{S} , we showed above that Fox's and Romoli's constraint do block this inference.

- (47) a. Some of the students read all of the books.
 b. $\not\rightarrow$ *some of the students didn't read any of the books*

On the other hand, if we recursively apply \mathcal{S} as in (48), the problem re-emerges.

$$(48) \quad \mathcal{S}(\mathcal{S}(\text{Some of the students read all of the books}))$$

The alternatives of the first application of \mathcal{S} are given below in (49) and the result of this first derivation is in (50).

$$(49) \quad \left\{ \begin{array}{l} \text{Some of the students read all of the books} \\ \text{All of the students read all of the books} \end{array} \right\}$$

$$(50) \quad \mathcal{S}(\text{Some of the students read all of the books}) =$$

The scalar inferences of negative quantifiers

some of the students read all of the books $\wedge \neg$ (all of the students read all of the books)

The alternatives over which the second application of \mathcal{S} in (48) operates are those in (51), which are in turn equivalent to (52). Again, given that the alternatives are in the scope of \mathcal{S} , the second alternative does not entail the third anymore, and the latter is thereby allowed by Fox's constraint.

$$(51) \quad \left\{ \begin{array}{l} \mathcal{S}(\text{Some of the students read all of the books}) \\ \mathcal{S}(\text{All of the students read all of the books}) \\ \mathcal{S}(\text{All of the students read some of the books}) \end{array} \right\}$$

$$(52) \quad \left\{ \begin{array}{l} \text{Some of the students read all of the books} \wedge \neg \text{all read all} \\ \text{All of the students read all of the books} \\ \text{All of the students read some of the books} \wedge \neg \text{all read all} \end{array} \right\}$$

The meaning of (48) given the alternatives in (52) is therefore (53), with the unwanted inference that some of the students didn't read any of the books.

$$(53) \quad \begin{aligned} &\mathcal{S}(\mathcal{S}(\text{Some of the students read all of the books})) = \\ &\text{Some of the students read all of the books} \wedge \neg(\text{all of the students read all} \\ &\text{of the books}) \wedge \neg(\text{all of the students read some of the books} \wedge \neg(\text{all of the} \\ &\text{students read all of the books})) = \\ &\text{Some of the students read all of the books} \wedge \neg(\text{all of the students read some} \\ &\text{of the books}) = \\ &\text{Some of the students read all of the books} \wedge (\text{some of the students didn't} \\ &\text{read any of the books}) \end{aligned}$$

8 Appendix B: Bar-Lev and Fox (2017)

In this Appendix, we sketch how Bar-Lev & Fox's (2016) derivation of the inferences in (54-a) and (54-b) from (54), repeated from above:

- (54) Every student took syntax or semantics.
- a. \rightsquigarrow *Some student took syntax*
 - b. \rightsquigarrow *Some student took semantics*

Their derivation of the inference in (54-a) and (54-b) involves a multiple scalar implicature computation on the sentence in (54) as indicated in (55).

$$(55) \quad \mathcal{S}_{A'}(\mathcal{S}_A(\text{every student took syntax or semantics}))$$

The first \mathcal{S} operates over the following set A of alternatives, none of which can be

excluded.³⁵

$$(56) \quad A = \left\{ \begin{array}{l} \text{Every student took Syntax,} \\ \text{Every student took Semantics,} \\ \text{Some student took Syntax,} \\ \text{Some student took Semantics} \end{array} \right\}$$

The second \mathcal{S} however now applies over the following set of alternatives A' which consists of the alternatives in (56) to which an implicature computation has been applied as in (57), the result of which is indicated in (58):

$$(57) \quad A' = \left\{ \begin{array}{l} \mathcal{S}_A(\text{Every student took Syntax}), \\ \mathcal{S}_A(\text{Every student took Semantics}), \\ \mathcal{S}_A(\text{Some student took Syntax}), \\ \mathcal{S}_A(\text{Some student took Semantics}) \end{array} \right\}$$

$$(58) \quad A' = \left\{ \begin{array}{l} \text{Every student took Syntax} \wedge \neg(\text{Some student took Semantics}), \\ \text{Every student took Semantics} \wedge \neg(\text{Some student took Syntax}), \\ \text{Some student took Syntax} \wedge \neg(\text{Some student took Semantics}), \\ \text{Some student took Semantics} \wedge \neg(\text{Some student took Syntax}), \end{array} \right\}$$

In (58), only the first two alternatives are excludable and their negation, together with the assertion, gives rise to the desired inferences that some of the students took Syntax and that some took Semantics.

$$(59) \quad \begin{array}{l} \text{Every student took Syntax or Semantics} \wedge \\ \neg(\text{Every student took Syntax} \wedge \neg(\text{some student took Semantics})) \wedge \\ \neg(\text{Every student took Semantics} \wedge \neg(\text{some student took Syntax})) = \\ \text{Every student took Syntax or Semantics and some student took Syntax and} \\ \text{some student took Semantics} \end{array}$$

9 Appendix C: Experimental items

³⁵ We are assuming that alternatives are derived following Romoli's (2012) algorithm. We are leaving out the alternatives involving disjunction and conjunction for simplicity, it is easy to show that they are not relevant for the inferences we are interested in here.

Every student ate some of the cookies
↪At least one of the students ate some of the cookies (every true)
Every student solved some of the problems
↪Not every student solved all of the problems (every weak)
Every student liked some of the games
↪No student liked all of the games (every strong)
Every student drove some of the cars
↪Not every student drove some of the cars (every false)
No student met all of the teachers
↪At least one of the student didn't meet all of the teachers (no true)
No student broke all of the glasses
↪Some student broke some of the glasses (no weak)
No student moved all of the chess pieces
↪All students moved some of the chess pieces (no strong)
No student scared all of the animals
↪Some student scared all of the animals (no false)

Table 3 Table A.1: Sentences and inferences in Experiment 1

The boys went to a toy fair and wanted to try out several radio-controlled cars.
Every one drove some of the cars
↪ Some one drove all of the the cars (compatible every)
The students of the Spanish language course wrote a vocabulary test.
Every one knew some of the words
↪ Some one knew all of the words (compatible every)
The friends stayed together at a wine tasting event where they could taste several wines.
Every one tasted some of the wine
↪Some one tasted all of the wines (compatible every)
The children were at a school competition. There were different types of games.
Every one participated in some of the games
↪Some one participated in all of the games (compatible every)
The children are at the farm and were allowed to pet the rabbits.
No one petted all of the rabbits
↪ Some one did not pet any of the rabbits (compatible no)
The florists each had several flowers at home, which had to be watered today.
No one watered all of the flowers
↪Some one did not water any of the flowers (compatible no)
The girls had riding lessons and they were given different horses to ride.
No one rode all of the horses
↪Some one did not ride any of the horses (compatible no)
The customers were shown as set of ear rings each.
No one bought all of the ear rings
↪Some one did not buy any of the earrings (compatible no)
The friends met in the evening for a board game party.
Every one liked some of the games
↪No one liked all of the games (strong every)
The children went to an animal shelter and they were allowed to feed the cats.
Every one fed some of the cats
↪No one fed all of the cats (strong every)
The children passed a candy shop. They were allowed to buy certain types of candies.
Every one bought some of the candies
↪No one bought all of the candies (strong every)
In preparation of the Carnival, the girls were given some dresses to try on.
Every one tried on some of the dresses
↪No one tried on all of the dresses (strong every)

Table 4 Table A.2: Sentences and inferences in Experiment 2

The students could choose between different classes at the beginning of term.
No one took all of the classes
↔Every one took some of the classes (strong no)
The visitors went to an exhibition which had several rooms.
No one liked all of the paintings
↔Every one liked some of the paintings (strong no)
The friends spent the evening in a karaoke bar. They all had a set of songs to sing.
No one sang all of the songs
↔Every one sang some of the songs (strong no)
The cleaning women had each a set of dishes to clean.
No one cleaned all of the dishes
↔Every one cleaned some of the dishes (strong no)
The children were each given a bag of cookies.
Every one ate some of the cookies
↔At least one of them ate some of the cookies (true every)
The students just had a series of tests.
Every one passed some of the exams
↔At least one of them passed some of the exams (true every)
The children are at a birthday party and they are allowed to watch a set of Disney movies.
Every one watched some of the movies
↔At least one of them watched some of the movies (true every)
The office staff was moving to new office rooms. Everyone was given a set of boxes.
Every one carried some of the boxes
↔At least one of them carried some of the boxes (true every)
The mothers were invited to open school day where they had the opportunity to meet different teachers.
No one met all of the teachers
↔At least one of them did not meet all of the teachers (true no)
The girls took an art class.
They has to copy a set of paintings.
No one copied all of the paintings
↔At least one of them did not copy all of the paintings (true no)
The friends went on a safari tour. They had a guide of the different types of animals.
No one saw all of the animals
↔At least one of them did not see all of the animals (true no)
At Easter, it's custom to hide eggs and let the children find them.
No one found all of the eggs
↔At least one of them did not find all of the eggs (true no)

Table 5 Table A.2: Sentences and inferences in Experiment 2, continued

<p>The students had a surprise exam. Every one solved some of the problems ~>Not every one solved all of the problems (weak every) The students had several historical dates to remember. Every one remembered some of the dates ~>Not every one remembered all of the dates (weak every) The pupils had English classes and were each given a set of poems. Every one read some of the poems ~>Not every one read all of the poems (weak every) The basketball players were at a tournament. Each had three shots to shoot. Every one hit some of the shots ~>Not every one hit all of the shots (weak every) The friends went to an optician to look for new glasses. They were each given a set of glasses. No one tried on all of the glasses ~>Some one tried on some of the glasses (weak no) The post men each had to deliver several parcels. No one delivered all of the parcels ~>Some one delivered some of the parcels (weak no) The boys took part in a quiz show. In the end, they each had to answer a series of difficult questions. No one knew all of the answers ~>Some one knew some of the answers (weak no) The girls had baked many brownies and each wanted to sell them on sports day. No one sold all of the brownies ~>Some one sold some of the brownies (weak no)</p>	<p></p>
---	---------

Table 6 Table A.2: Sentences and inferences in Experiment 2, continued
