Is there *any* licensing in non-DE contexts? An experimental study[1]

Stavroula ALEXANDROPOULOU — *Utrecht University*
Lisa BYLININA — *Leiden University*
Rick NOUWEN — *Utrecht University*

Abstract. Why is weak NPI *any* sometimes, but not always, licensed in non-downward entailing environments? In this paper, we present a series of experiments, where we probe the role of context in licensing *any* in the scope of various quantifiers. We compare our results to predictions made by three theories of exceptional NPI licensing. We show that contextual reasoning plays a role in non-monotonic environments, but that it does not in the scope of upward entailing quantifiers. Surprisingly, our results also show individual differences between the non-monotonic environments created by different quantifiers.

Keywords: polarity, monotonicity, NPIs, quantification, numerals.

1. Introduction

There is a long tradition in the literature of claims that downward entailingness is an important ingredient in accounting for the distribution of negative polarity items (NPIs) (Fauconnier, 1975; Ladusaw, 1979). An example like (1) is felicitous since the NPI *anything* is in an environment that is downward entailing (DE) (viz. the scope of *none of the boxes*). In contrast, (2) is infelicitous since *some of the boxes* does not create such an environment.

(1) None of the boxes contain anything.

(2) *Some of the boxes contain anything.

There is consensus, however, that downward entailingness is neither a sufficient nor a necessary condition for NPI licensing. Here, we focus on the latter. Following Linebarger (1987), it is widely acknowledged that not all occurrences of NPIs are occurrences in DE environments. Among the well-known exceptions are non-monotonic quantifiers like *exactly*, see example in (3).

(3) Exactly two of the boxes contain anything.

These kinds of licit occurences of NPIs in non-DE environments are not general, however. Minimally different examples like (4) are judged less acceptable, for instance.

(4) ??Exactly 98 of the boxes contain anything.

Our focus in this paper is to come to a better understanding why weak NPIs like the kind exemplified by *any* and its kin (wNPIs) are licensed in some, but not all, non-DE environments. In particular, we focus on different proposals in the literature as to why cases like (3) exist.[2] These

[1]We would like to thank Chris Barker, Mingya Liu, Stephanie Solt, and Yasu Sudo for their constructive feedback, as well as the audiences of XPrag 2019, SuB24, ESPP 2019, and of the workshop 'Semantics in Athens III'.

[2]These proposals, as is our research, are restricted to *any*-style NPIs. This is because only this subclass of polarity sensitive expressions is licit in (some) non-DE environments. That is, (i), with the strong NPI *in years*, contrasts starkly with (3).

(i) *Exactly two of the boxes have been opened in years.

 In: M. Franke et al. (eds.) *Proceedings of Sinn und Bedeutung 24*, vol. 1, pp.35–47. Osnabrück University.

theories share the idea that wNPIs are licit in a non-DE environment only under specific contextual circumstances. They differ in the properties of the relevant contextual circumstances, giving rise to different predictions.

We conducted two experiments that test these predictions. Our first experiment tests the acceptability of *any* in environments with different entailment properties. In our second experiment, we investigate to what extent readers of sentences containing wNPIs in non-DE environments make assumptions about the context.

The plan is as follows: We will first briefly introduce three accounts of sentences like (3) and extract testable predictions. We then present our experiments and end by discussing what theoretical conclusions we can draw from our results as well as by suggesting some ideas for further investigation.

2. Theoretical background

We will compare three main approaches to NPI licensing in non-DE environments.

Under one of these, cases like (3) are seen as exceptional in that they are not actual cases of NPI licensing. Giannakidou (2008) proposes that NPIs are only licensed in *non-veridical environments*, see (5) for the definition of a non-veridical environment (= the scope of a non-veridical operator).

(5) *(Non-)veridicality of propositional operators*
A propositional operator O is veridical iff $O(p)$ entails or presupposes that p is true in some individual's epistemic model $M_E(x)$; otherwise O is non-veridical.

Ignoring the epistemic model part, (5) says that the scope of an operator O is non-veridical if the truth of p does not follow from $O(p)$. Beyond this theory of licensing, Giannakidou proposes the existence of a *rescuing mechanism* that accounts for licit occurrences of NPIs in veridical environments.

Quantifiers are not propositional operators, so it is not trivial to apply (5) to examples like (3) or (4). We assume that what matters for veridicality of quantifiers is the existence of a witness. For instance, 'exactly n students' is veridical for $n > 0$ since 'exactly n students passed the test' entails that there were students that passed the test. Under this generalisation of (non)veridicality to non-propositional cases, *anything* in (3) is an example of an NPI occurrence in a veridical environment. As such, on Giannakidou's theory, the NPI is not *licensed*, but rather *rescued*. In order for a wNPI to be rescued, there has to be some contextually available parallel sentence that *would* license the NPI. For the case of (3) this would for instance be a statement that the number of non-empty boxes was low:

(6) Not many boxes had anything in them.

Not many creates a non-veridical environment, since (6), for example, is compatible with a situation where all boxes were empty. Rescuing is less likely for the sentence in (4) since it is harder to connect to a parallel sentence with a non-veridical environment like (6). Overall, the rescuing mechanism does not distinguish between non-monotone and upward-entailing environments in a systematic way.

Diametrically opposed to this theory is that by Crnič (2014), where cases like (3) are analysed

in terms of a proper licensing mechanism. Crnič argues that wNPIs such as *any* trigger scalar alternatives, roughly: *anything* $\supset$ *two things* $\supset$ *three things* etc. In terms of the alternatives that *anything* triggers, it is semantically identical to an indefinite with covert numeral *one* – its alternatives involve higher quantities.

Such weak scalar items can be licensed by a covert *even*: that is, wNPIs in non-monotone environments are associates of covert *even*. The silent *even* is a covert counterpart of the overt focus particle *even* – a clause-level propositional operator that requires its propositional argument, its prejacent, to be less likely than all the relevant focus alternatives to the constituent it is adjoined to (scalar presupposition; Karttunen and Peters, 1979; Wilkinson, 1996, and many others). The likelihood relation is provided by the context.

Here is a felicitous use of overt *even* in a context where "Syntactic Structures" is the least likely book to be read, out of some contextually relevant set of books (LF and interpretation in (7b) and (7c), respectively):

(7) a. John read even SYNTACTIC STRUCTURES.
b. [even$_C$ [John read [Syntactic Structures]$_F$]]
c. $[\![(7a)]\!]^c$ is defined only if for all relevant q in {that John read x: x is a book}: that John read Syntactic Structures $< q$.
If defined, $[\![(7a)]\!]^c = 1$ iff John read Syntactic Structures in w.

(adapted from Crnič (2014))

In a similar way, the felicitous use of *anything* in (3) relies on the way probabilities are set up by the context. (3) is acceptable if it is less likely that exactly two boxes have one thing in them than that exactly two boxes have two things in them, and so on. As Crnič shows, this is the case whenever there is the conditional expectation that more boxes have something in them.

One example of probabilities that the context might contain that would license *anything* in (3) is in Figure 1 (see next page). If we are interested in two boxes and how many things they are expected to contain, 1 is the least likely alternative. With the number of boxes increasing, the relations between probabilities also change.

As such, Crnič (2014) can account for the contrast between (3) and (4) on intuitively similar grounds as does Giannakidou, even though the mechanism is somewhat different. (4) is bad according to Crnič because it makes it unlikely that the speaker considers the number of boxes that contain stuff low.

Finally, Barker (2018) proposes that wNPIs are *scope licensed*: they are items that signal they have narrow scope relative to some other operator. Barker argues that signalling narrow scope is only useful when the wide scope interpretation does not entail the narrow scope interpretation – that is, when it is not the case that the narrow scope interpretation is the less informative one.

One environment where the wide scope interpretation of an existential entails the narrow scope interpretation is the scope of *every*, see (8). The wide scope interpretation (a single book that is read by every woman) entails the narrow scope interpretation.

(8) Every woman read [a book].
$\forall > \exists$ entails $\exists > \forall$

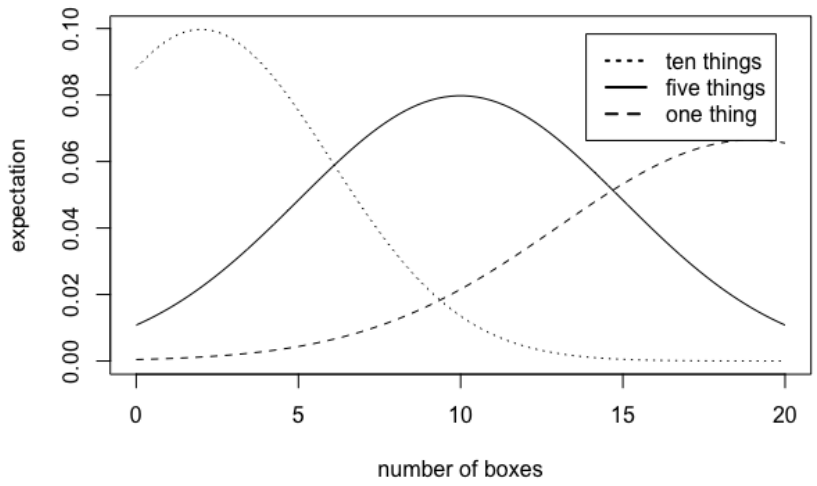

Figure 1: Contextual probabilities licensing *anything* in (3), according to Crnič (2014).

Barker capitalizes on the parallel between such entailments and the lack of NPI licensing:

(9) *Every woman read [any book].

The consequence of the scope licensing view is that wNPIs are licensed only in non-upward entailing contexts:

(10) (Barker, 2018)
An NPI is scope licensed in a context only if a wide scope existential binding a variable in the position of the NPI does not entail a narrow scope existential binding that position.

Barker further assumes that scope-licensing needs to be supplemented with contextual constraints. That is, for Barker (2018) scope licensing is a necessary yet not a sufficient constraint, but he does not commit to any specific set of such constraints.

Here are the predictions derived from these three proposals that we test in two experiments:

Giannakidou (2008) Licit use of a wNPI in a non-DE environment (i.e., *rescuing*) is in principle possible in any veridical environment. That is, rescuing occurs both in the scope of upward monotone and non-monotone operators, as long as there is suitable contextual pressure.

Crnič (2014) Licit use of a wNPI in a non-DE environment is only possible in the scope of non-monotone operators (because the semantic condition on *even*-licensing cannot be fulfilled in an upward entailing context), always subject to suitable contextual pressure.

Barker (2018) Licit use of a wNPI in a non-DE environment is only possible in the scope of non-monotone operators. There are no specific predictions with respect to the role of context.

In what follows, we present our experiments (Experiment 1 and Experiment 2) that aimed at testing the above predictions.

3. Experiments

Experiment 1 aimed to test to what extent wNPIs, as represented by *any*, are accepted in the scope of non-downward entailing quantifiers. Experiment 2 tests whether the occurrence of wNPIs in non-downward entailing environments is modulated by a contextually inferred expectation.

3.1. Experiment 1

Experiment 1 consisted of an acceptability judgement task, carried out in English.

3.1.1. Methods

Participants. We recruited 39 participants via Amazon Mechanical Turk and excluded one of them because their native language was other than English. The data of 32 native English participants were included in the subsequent analysis, as we removed the data of 6 participants who judged correctly fewer than 75% of the filler items. All participants received $1.40 for participation in the study.

Materials & procedure. We showed participants sentences like (11), where QUANT stands for one of the following quantifiers: *at least n*, *at most n*, *exactly n* and *between n and m*[3] (conditions ATLEAST, ATMOST, EXACTLY, BETWEEN, respectively), and the sentence either contained a DP headed by the wNPI *any* or the corresponding bare plural (factor POL; conditions NPI and BARE, respectively). Hence, the experiment had a 4 × 2 design, that is, experimental items appeared in 8 conditions.

(11) QUANT products had $(\text{ANY})_{Pol}$ artificial sweeteners in them.

Is this an acceptable sentence of English?
(click on your answer)

- Yes
- No

We asked participants to indicate whether sentences like (11) were acceptable sentences of English by clicking on *Yes* or *No*. We tested 16 experimental items intermixed with 32 filler items, 16 of which were designed so as to evoke a NO response while the rest to evoke a YES response. Each participant saw all eight conditions and two experimental items per condition, as well as the same 32 fillers. The total of 48 items was randomly ordered for each participant.

Participants were first asked to give their consent to participate in the study. Those doing so proceeded to the instructions of the experiment and after that to the main part of the experiment.

[3] Across the different items, the choices for *n* and *m* were always small numbers.

3.1.2. Predictions

Under the valid assumption that responses arising from pragmatic reasoning are less readily given than responses where no such reasoning is required (Bott and Noveck, 2004; Cummins and Katsos, 2010; Katsos and Bishop, 2011, a.o.), Giannakidou's (2008) approach predicts a difference in acceptability between ATMOST (non-veridical) and the other QUANT conditions with *any*, which all include veridical environments. ATMOST is the only condition where the wNPI does not need *rescuing* by means of pragmatic reasoning, but is licensed by a proper grammatical mechanism (non-veridicality).

On the contrary, according to Crnič (2014), one expects an additional[4] discrepancy in acceptability between ATLEAST on the one hand, and EXACTLY and BETWEEN conditions on the other hand. On this account, only non-upward entailing environments allow wNPI licensing by covert *even*, hence the upward-entailing ATLEAST condition is expected to receive low acceptance rates in the NPI condition – lower than any of the three corresponding non-upward entailing QUANT conditions.

Lastly, on Barker's account too, wNPIs are licensed only in non-upward entailing contexts (scope licensing). Thus, *any* in the ATLEAST condition is predicted to receive very low acceptance rates as compared to the ATMOST condition. However, in the absence of further explication, the non-monotone QUANT conditions could go either way, depending on possible further (contextual) licensing conditions on top of scope licensing.

3.1.3. Results & discussion

The collected categorical data (after the removal of bad subjects) were analysed with binomial generalised mixed models (glmer) using the `lme4` package (Bates et al., 2014) in `R` (R Core Team 2019). Figure 2 shows the proportion of YES responses per experimental condition.

We set ATMOST and BARE as the reference levels of the factors QUANT and POL, respectively. Via model comparison, the model including the interaction term turned out to have the best fit as compared to the model with main effects for QUANT and POL ($\chi^2(3) = 13.533$, $p < .01$). Our model also included intercept random effects for participants and items.

As indicated by the lower acceptance rates of the NPI condition in Figure 2, the addition of the wNPI *any* to the experimental sentences reduces their acceptability overall. Indeed, our statistical analysis specifically revealed that adding the wNPI *any* to the AT MOST items significantly reduces their acceptability ($SE = .687$, $z = -4.817$, $p < .0001$). This is quite surprising, as *at most* was taken to be a paradigmatic NPI-licensing environment and chosen for this reason as the baseline condition of QUANT. However, this finding seems to be in line with Sanford et al.'s (2007) finding that *at most* might not be that negative after all, who also demonstrated

[4]Here too, the licit occurrences of wNPIs in non-monotone environments are context-dependent to some extent. Satisfying the scalar presupposition associated with covert *even* in non-monotone environments is contextually modulated by a commonly shared expectation. Such contextual modulation does not apply in the case of downward entailing contexts (any DE context satisfies the probability requirement). Thus, wNPI *any* will be less acceptable in the non-monotone EXACTLY and BETWEEN conditions than in the downward entailing ATMOST condition – under the assumption that contextual reasoning is reflected in acceptability rates.

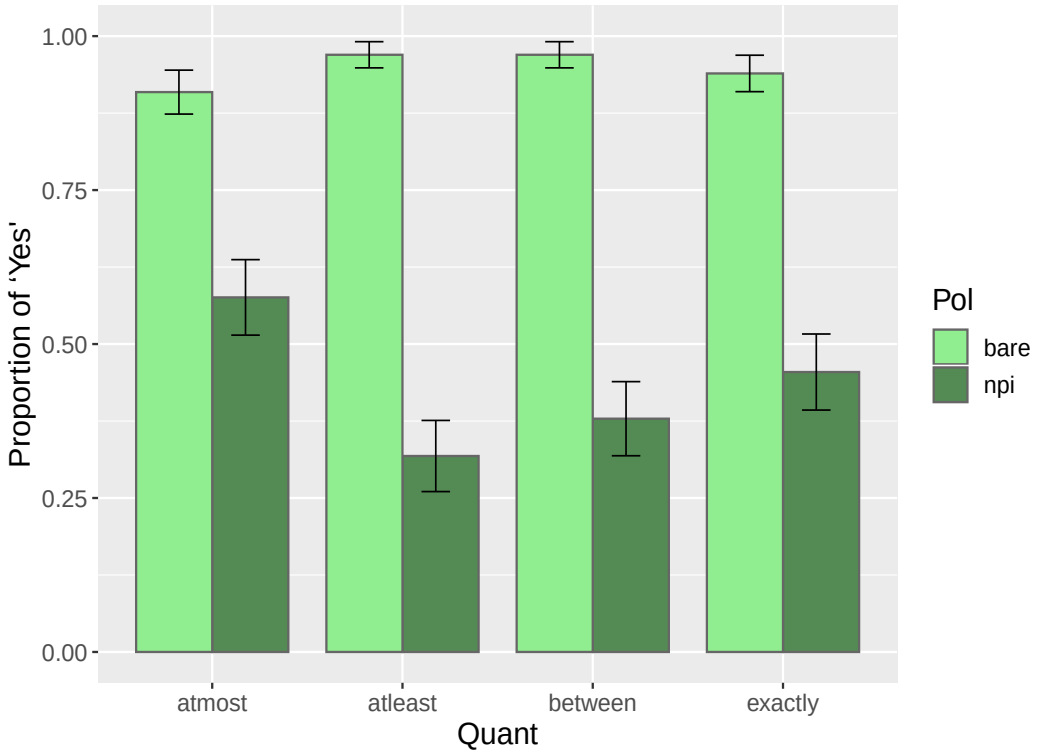

Figure 2: Mean response proportions and SEs per condition in the acceptability judgement task.

that *at most* is actually less negative than the downward entailing quantifiers *none*, *not many*, *few*, and *less than n*. We come back to this in section 4.

The statistical analysis further showed that participants judged the difference in acceptability between BARE and NPI conditions to be significantly greater for ATLEAST and for BETWEEN than for ATMOST (respective interaction effects: NPI×ATLEAST: $SE = 1.074$, $z = -3.128$, $p < .01$; NPI×BETWEEN: $SE = 1.063$, $z = -2.712$, $p < .01$). However, this effect did not reach significance for EXACTLY (NPI×EXACTLY: $p = .102$). That is, the acceptability drop that the addition of NPI causes is larger for ATLEAST and for BETWEEN , but not for EXACTLY, as compared to ATMOST.

The above results suggest that there is a discrepancy among the three non-downward entailing quantifiers in terms of acceptability (in the presence) of the wNPI *any* in their scope as compared to *at most*. More specifically, it is the class of non-monotone quantifiers that seems not to behave in a homogeneous way in that respect. This is not expected according to Giannakidou's and Crnič's accounts, while it appears to be consistent with the variable predictions that Barker's account derives as to the non-monotone quantifiers.

In the next section, we report on the results of Experiment 2, which aims to uncover whether contextual expectation plays a role in rendering the uses of wNPs in different non-downward entailing environments licit and acceptable.

3.2. Experiment 2

Experiment 2 too was an offline judgement task conducted in English.

3.2.1. Methods

Participants. 56 native speakers of English participated in this experiment administered on Amazon Mechanical Turk. We rejected the data of 17 of them who answered correctly fewer than 75% of the filler items considered for participant removal. Thus, in the final analysis we included the data of the remaining 39 participants. All participants received $1.60 for taking part in this experiment.

Materials & procedure. In this experiment, we presented participants with the same items as in Experiment 1, which were all preceded by the sentence *I didn't expect this*, see (12). The design was exactly the same as in Experiment 1, i.e., QUANT: ATLEAST, ATMOST, EXACTLY, and BETWEEN, and POL: NPI and BARE.

(12) I didn't expect this, but QUANT products had $(\text{ANY})_{Pol}$ artificial sweeteners in them.

What do you think the writer of the sentence expected?
(click on your answer)

- that more products had artificial sweeteners in them
- that fewer products had artificial sweeteners in them

Participants now had to decide what the writer of the sentence expected: that is, illustrating for (12), whether they expected more products to have artificial sweeteners (higher expectation: HIEXP) or fewer (lower expectation: LOEXP). Our 16 experimental items were interspersed with 32 fillers, 10 of which contained an explicit bias toward a higher expectation on the part of the writer (*higher bias*; see (13)), 10 of them created a lower-expectation bias (*lower bias*; see (14)) and the remaining 12 were unbiased, see (15).

(13) I didn't expect this, but only 2 of the invited speakers of the conference were female.

(14) I didn't expect this, but the parcel arrived after 10:30.

(15) I didn't expect this, but 5 climbers reached the summit.

Like Experiment 1, each participant saw all eight conditions and two experimental items per condition, as well as the same 32 fillers. The total of 48 items was randomly ordered for each participant.

Lastly, the participants who gave their consent to participate in the study were presented with the instructions and, before they proceeded to the main part of the experiment, they were also given two practice items in order to familiarise themselves with the task.

3.2.2. Predictions

For Giannakidou (2008), contextual licensing (i.e., *rescuing*) is a general option, available in any veridical environment, that is, it can apply in both upward entailing and non-monotone en-

vironments. To illustrate for the item in (12),[5] sentence (16) would be made contextually available via the rescuing mechanism, licensing *indirectly* the wNPI *any* in the ATLEAST (upward entailing) and in the two non-monotone QUANT conditions of (12), BETWEEN and EXACTLY, due to the presence of the non-veridical operator *not*.

(16) Not many products had artificial sweeteners in them.

Given that sentence (16) may imply a higher prior expectation on the part of the writer as to the number of products with artificial sweeteners (see Sanford et al., 2007 on *not many*), the NPI condition of all three veridical quantifiers (*at least*, *between*, *exactly*) is expected to receive more HIEXP responses than their corresponding *bare* condition, and particularly more than in the case of ATMOST. This is because the wNPI *any* in the ATMOST condition is already (grammatically) licensed by non-veridicality, hence no pragmatic reasoning (relating to a higher expectation) applies.

On Crnič's account, wNPIs are properly licensed by covert *even* (only) in non-upward entailing contexts and specifically when its scalar presupposition is satisfied. That is, the ATMOST, BETWEEN, and EXACTLY version of (12) with *any* is felicitous if it is less likely that QUANT products had one artificial sweetener in them than that QUANT products had two artificial sweeteners, etc. (see footnote 5). Unlike downward entailing quantifiers, for non-monotone quantifiers this is the case whenever there is the conditional expectation that more products had some artificial sweetener in them. Given that, it is predicted that (higher) expectation plays a role only in non-monotone contexts and that the difference in HIEXP responses between the NPI and the BARE conditions will be greater for the two non-monotone QUANT conditions than for the ATMOST condition. Contextual expectation as discussed in this account will not play any role in the ATLEAST condition – in such contexts wNPIs are not licensed to begin with as the presupposition of covert *even* is unsatisfiable.

Finally, from Barker's theory, it follows that upward entailing contexts cannot host wNPIs. In that sense, the predictions of his account are similar to Crnič's account. However, since Barker takes scope licensing only to be a necessary condition, there could be variation in the extent to which scope-licensing quantifiers need contextual reasoning to allow the NPI. Hence, while contextual expectation is not predicted to play any role in the ATLEAST items with *any*, we cannot derive any specific predictions as to the relative proportion of HIEXP responses to the three non-upward entailing QUANT conditions with *any*.

3.2.3. Results & discussion

The collected and cleaned categorical data on the basis of the higher- and lower-bias filler items were analysed with mixed-effects logistic regression models, similarly to Experiment 1. Our analysis included again QUANT and POL as predictors, with ATMOST and BARE as the respective references levels, and had intercept random effects for participants and items. Figure 3 shows the proportion of HIEXP responses per experimental condition.

Including the interaction term to the model significantly improved the model fit compared to the model with simple main effects for QUANT and POL ($\chi^2(3) = 9.748$, $p < .05$).

[5]Note that, as in experiment 1, the number in the QUANT conditions was always small.

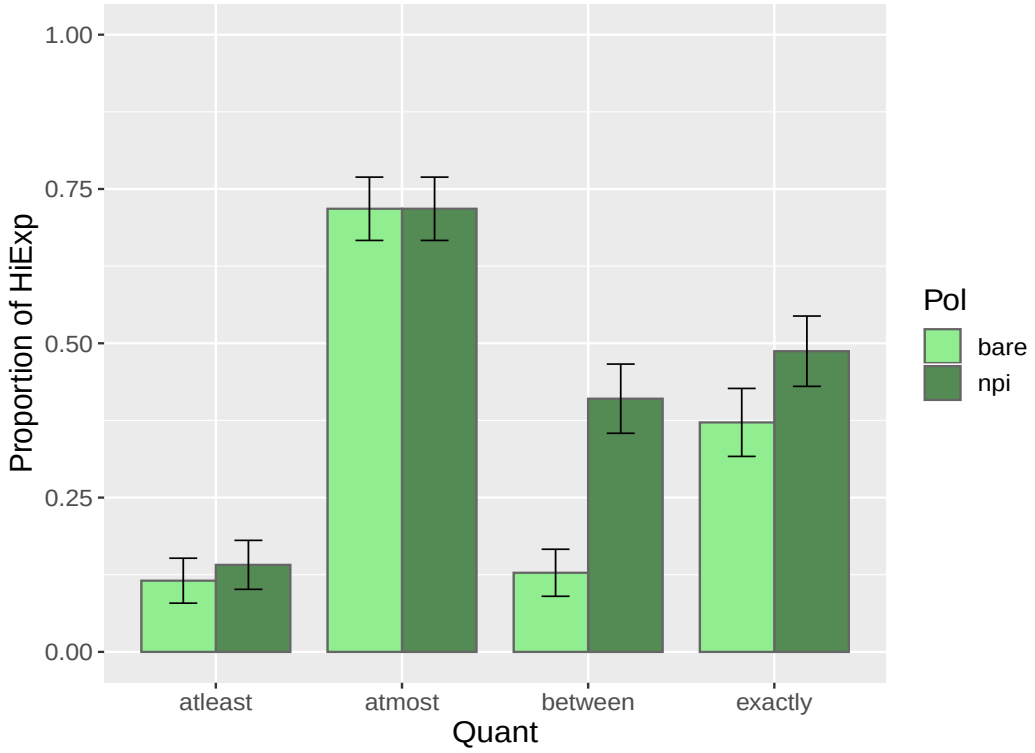

Figure 3: Mean response proportions and SEs per condition in the contextual expectation experiment.

As Figure 3 displays, HIEXP seems to play no role in the ATLEAST items, while the opposite appears to be the case for the ATMOST condition overall. In particular, the statistical analysis showed that the ATMOST condition received reliably more HIEXP response rates at the baseline level of the BARE POL condition compared to all three non-downward entailing QUANT conditions (ATLEAST: $SE = .451$, $z = 6.949$, $p < .0001$; BETWEEN: $SE = .446$, $z = 6.872$, $p < .0001$, and EXACTLY: $SE = .364$, $z = 4.359$, $p < .0001$). This reveals that the use of the quantifier *at most* is regulated by a contextually available (higher) expectation (i.e., of an amount that is more than the asserted one). Interestingly, this turns out to be the case regardless of the addition of the wNPI *any* (simple effect of NPI: $p = .977$) and, thus, is in line with the claim that the use of negative quantifiers signals that the speaker had expected a bigger amount than that asserted (Sanford et al., 2007, and references therein).

More importantly, our analysis further showed that the difference between BARE and NPI was significantly larger for BETWEEN than for ATMOST (interaction effect NPI×BETWEEN: $SE = .563$, $z = -2.891$, $p < .01$), but this effect was not significant for the other non-monotone QUANT condition (NPI×EXACTLY: $p = .328$) or the upward-entailing condition ATLEAST (NPI×ATLEAST: $p = .733$). Hence, of all three non-downward entailing quantifiers tested, the presence of a wNPI triggers contextual reasoning relating to a higher expectation only in the scope of *between*, resulting in a non-homogeneous behaviour of the class of non-monotone quantifiers, similar to that found in Experiment 1.

The above results are at odds with Giannakidou's proposal, which would have us expect similar

interaction effects for the veridical quantifiers *at least* and *exactly* too. That is, we only observe a role of context in one non-monotone environment and not in an upward entailing one. The results are also at odds with Crnič's predictions since in his theory it would be expected that all non-monotone quantifiers, including *exactly*, display an interaction effect similar to that of *between*. The results are in line with Barker's theory, but only because this theory allows idiosyncratic conditions additional to scope licensing and could posit that *exactly* does not need contextual reasoning to license wNPIs, but *between* does.

4. General discussion

We studied the acceptability and the triggering of contextual reasoning of sentences containing four different quantifiers with and without a weak NPI in their scope. Surprisingly, the two non-monotonic quantifiers behave differently. The acceptability of a wNPI in the scope of *exactly* is comparable to that of a wNPI in the scope of *at most*. Similarly, as is the case for *at most*, the presence of a wNPI in the scope of *exactly* has limited impact on the extent to which contextual reasoning is triggered. In contrast, the acceptability of a wNPI in the scope of *between* is more comparable to what we observed for *at least*. Moreover, *between* was the only quantifier for which we observed a significant impact of the presence of a wNPI on contextual reasoning.

The fact that the non-monotone quantifiers as a whole do not perform in a uniform way goes against both Giannakidou's and Crnič's theories. On the other hand, it is in line with Barker's proposal, though not in a particularly enlightening way. The theory in Barker (2018) leaves space for differences between different non-monotone quantifiers in terms of NPI licensing, non-vacuous scope marking being a necessary but not a sufficient condition for NPI licensing.

So why do we observe the above discrepancy between *between* and *exactly*? It could be, of course, that our experiments were simply not sensitive enough to pick up effects for *exactly* quantifiers, similar to those for *between*. A more sensitive method or perhaps changing from a binary response to a Likert scale could help clarify the status of *exactly*.

There is, however, also a very real option that our results are indicative of some fundamental interpretative difference between *exactly* and *between*. One possibility is that our relevant finding could have to do with the so-called 'phantom' *at least* readings that *between n and m* quantifiers have been claimed to have.

According to Marty et al. (2015), sentences with *between* are ambiguous between an 'exactly' reading and an 'at least' reading. Say, (17), under the 'exactly' reading would be true if four students were late – and false if six (or any number higher than five) students were late. This reading is detectable intuitively and perceived as the only reading of (17). However, Marty et al. (2015) argue that (17) has one more reading, namely, an 'at least' reading, under which (17) is true if six students (or any other number of students higher than three) were late:

(17) Between three and five students were late.

This reading is argued to be ruled out on pragmatic grounds, and thus, according to Marty et al. (2015), *between*-sentences are never used to convey the 'at least' reading – therefore the term 'phantom' reading. Still, some theories predict that both readings of *between*-sentences are available and are generated by grammar. Marty et al. (2015) in fact show in a series of experiments that both an 'exactly' and an 'at least' reading are available with *between*-sentences.

Importantly, the availability of 'at least' readings sets *between* apart from *exactly*. As the ambiguity in question affects the monotonicity profile of *between*, one might expect this ambiguity to affect the NPI licensing properties of *between* as well. This effect would be in the direction of decreased acceptability of NPIs in the scope of *between* due to the interference of the *at least* reading of *between* – as well as of a more important role of contextual factors as means of disambiguation. This is exactly what we observe in our experiments.

One way to pin down the role of 'phantom' readings of *between* in our study would be to control for factors that give rise to such readings. In particular, these readings only arise in distributive, as opposed to collective, contexts. As an idea for further investigation, one could compare the NPI-licensing behaviour of *between* in distributive contexts to that in collective contexts.

As a final remark, another idea of a follow up study would be to try out the same experiments with a different baseline. As noted in the previous section, on the assumption that the downward entailing quantifier *at most* creates a paradigmatic environment for NPI licensing, we took ATMOST to be the baseline condition for QUANT in our two experiments. However, this turned out to receive unexpectedly low acceptance rates when *any* was present (58% 'Yes' in Experiment 1). This finding is in line with Sanford et al.'s (2007) finding that *at most* is not a very negative quantifier, while *less than* has been found to be considerably more negative. Based on this finding, in order to have a more accurate baseline quantifier condition, it would be instructive to replace the superlative quantifiers *at most* and *at least* in our experiments with the corresponding comparative quantifiers *less/fewer than* and *more than*.

References

Barker, C. (2018). Negative polarity as scope marking. *Linguistics and Philosophy 41*(5), 483–510.

Bates, D., M. Maechler, B. Bolker, and S. Walker (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Bott, L. and I. Noveck (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language 51*(3), 437–457.

Crnič, L. (2014). Non-monotonicity in NPI licensing. *Natural Language Semantics 22*(2), 169–217.

Cummins, C. and N. Katsos (2010). Comparative and superlative quantifiers: Pragmatic effects of comparison type. *Journal of Semantics 27*(3), 271–305.

Fauconnier, G. (1975). Polarity and the scale principle. In *Papers from the 11th regional meeting of the Chicago Linguistic Society*.

Giannakidou, A. (2008). Negative and positive polarity items: Variation, licensing, and compositionality. In K. von Heusinger, C. Maienborn, and P. Portner (Eds.), *Semantics, An international handbook of meaning*. In preparation.

Karttunen, L. and S. Peters (1979). Conventional implicature. *Syntax and Semantics 11*, 1–56.

Katsos, N. and D. V. Bishop (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition 120*(1), 67–81.

Ladusaw, W. (1979). *Polarity sensitivity as inherent scope relations*. Ph. D. thesis, The University of Texas at Austin.

Linebarger, M. C. (1987). Negative polarity and grammatical representation. *Linguistics and Philosophy 10*, 325–387.

Marty, P., E. Chemla, and B. Spector (2015). Phantom readings: the case of modified numerals. *Language, Cognition and Neuroscience 30*(4), 462–477.

R Core Team (2019). R: A language and environment for statistical computing.

Sanford, A. J., E. J. Dawydiak, and L. M. Moxey (2007). A unified account of quantifer perspective effects in discourse. *Discourse Processes 44*(1), 1–32.

Wilkinson, K. (1996). The scope of *even*. *Natural Language Semantics 4*, 193–215.