The influence of polarity items on inferential judgments*

Milica Denić, Vincent Homer, Daniel Rothschild, Emmanuel Chemla March 19, 2019

Abstract

Polarity sensitive items are linguistic expressions such as *any*, *at all*, *some*, which are felicitous in some linguistic environments but not others. Crucially, whether a polarity item is felicitous in a given environment is argued to depend on the inferences (in the reasoning sense) that this environment allows. We show that the inferential judgments reported for a given environment are modified in the presence of polarity sensitive items. Hence, there is a two-way influence between linguistic and reasoning abilities: the linguistic acceptability of polarity items is dependent on reasoning facts and, conversely, reasoning judgments can be altered by the mere addition of seemingly innocuous polarity items.

Keywords: modularity; polarity; monotonicity; intuitions; reasoning.

1 Polarity and monotonicity

Monotonicity is an abstract, logical property that a linguistic environment is said to have when this environment systematically supports inferences from subsets to supersets (or *vice versa*). For instance, an environment is *upward monotone* or *upward-entailing* if it supports a subset to superset inference; an example is the environment of the boldface expressions in (1). Similarly, a *downward monotone* or *downward-entailing* environment supports the superset to subset inference; an example is in (2).

(1) This animal is **a siamese cat**.
$\rightsquigarrow$ This animal is **a cat**.

(2) This animal isn't **a cat**.
$\rightsquigarrow$ This animal isn't **a siamese cat**.

Interestingly, there is a class of expressions, called *polarity items* (PIs) whose acceptability has been linked to this logical property of monotonicity. This was first proposed by Fauconnier (1975) and Ladusaw (1979), in relation to the most studied category of such expressions, namely negative polarity items (NPIs) such as *any*, *ever*, and *at all*. The generalization proposed for the distribution of NPIs is that they are felicitous in a downward-entailing environment, as in (3), and not felicitous in an upward-entailing environment, as in (4).

*The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and was supported by ANR-17-EURE-0017 and by a travel grant from the ESF as part of the Euro-Xprag network.

(3) This animal isn't a cat **at all**.

(4) *This animal is a cat **at all**.

In addition, there are *positive polarity items* (PPIs) such as *some*, *something*, *someone* that are acceptable in upward-entailing environments as in (5), but that cannot be interpreted in a number of downward-entailing environments. For instance, (6) doesn't have a reading in which the existential quantifier *some coffee* is interpreted with narrow scope with respect to negation (i.e. a reading equivalent to 'I didn't drink any coffee'). It is the lack of narrow scope of *some* under negation which signals that it is a PPI. Note that the sentence is acceptable under a reading where *some* outscopes negation (i.e. a reading equivalent to 'There is some coffee that I didn't drink'), possibly by moving covertly past negation, and therefore ending up being interpreted in an upward-entailing environment.

(5) I drank some coffee.

(6) I didn't drink some coffee. ($\neq$ I didn't drink any coffee.)

There is a large and complex literature seeking to refine and explain the exact conditions under which PIs are licit. Be that as it may, everyone agrees that there is some connection between the acceptability of NPIs and PPIs and the logical property of monotonicity: some authors argue that PIs are essentially connected with monotonicity, other may argue that monotonicity is not essential to the licensing constraints, and that it is, in the worst case, an accidental correlation. The former approach typically seeks reasons coming from the meaning of PIs that make them unacceptable outside of environments with certain monotonicity properties (Chierchia 2013, Kadmon and Landman 1993, Krifka 1991, 1995, Lahiri 1998). The latter approach, instead, proposes that PI licensing is a form of syntactic agreement with some operators (e.g., negations), and that these operators happen to induce an environment of a particular monotonicity (see Guerzoni 2006, Herburger and Mauck 2007, Szabolcsi 2004, Progovac 2000, among others). Crucially though, all must recognize that this correlation between the monotonicity properties of the environment and the PI licensing exists. Even if it is an accidental correlation, it could be taken advantage of: PIs may serve as a signal of monotonicity, and their mere presence at the surface level could (even if probabilistically) facilitate corresponding monotonicity inferences.

In this paper we explore the relationship between polarity items and reasoning capacities. In particular, we explore the question of whether the presence of a polarity item influences what inferences subjects are willing to make. Szabolcsi et al. (2008) investigated this question before in relation to negative polarity items, concluding that there is no influence of NPIs on downward inferences. Given that there is at the very least a probabilistic link between PIs and monotonicity, it is surprising not to find such an effect. In the experiments reported below, we find such an effect in various cases, albeit not the cases tested by Szabolcsi et al., for reasons we will discuss. Thus, we demonstrate that the presence or absence of PIs in a sentence *does* influence which inferences subjects are willing to make, thereby demonstrating that high level reasoning tasks can be influenced by what otherwise looks like innocent linguistic decorations.

2 Previous results

Psycholinguistic studies have investigated the licensing of polarity items using a variety of tasks including acceptability judgments (e.g., Drenhaus et al. 2005, Muller and Phillips 2018), ERP measures (e.g., Drenhaus et al. 2006, Drenhaus, Heiner et al. 2007, Saddy et al. 2004, Steinhauer et al. 2010, Shao and Neville 1998, Yurchenko et al. 2013, Yanilmaz and Drury 2018, Xiang et al. 2009), self-paced reading (e.g., Parker and Phillips 2016, Xiang et al. 2013) and eye-tracking (e.g., Vasishth et al. 2008). Here we focus on two studies which jointly investigated the licensing of polarity items and inferential judgments of monotonicity: Chemla et al. (2011) and Szabolcsi et al. (2008).

Chemla et al. (2011) collected from a group of people both upward/downward inferential judgments and NPI acceptability judgments: it was found that the inferences a particular person considers valid in a given linguistic environment predict how acceptable they would find an NPI in that same environment. This study thus provided empirical confirmation of the relationship between monotonicity properties and NPI acceptability. In fact, these results also suggest that *subjective* individual judgments of inferential properties are a better indicator of PI acceptability than objective, logical upward and downward-entailingness. As in the experiments below, the study did not test all-or-nothing judgments of either NPI-acceptability or monotonicity inferences, but rather looked at graded judgments. The generalization reached about the determinants of NPI acceptability were more 'graded' than those in the syntax/semantics literature. In particular, they found that downward-entailingness and upward-entailingness *together* were a better predictor of NPI acceptability than either alone was, NPIs are thus good in environments to the extent that they are perceived as DE and/or as not-UE.

The other relevant study on the connection between PI acceptability in an environment and the monotonicity properties of that environment is the aforementioned Szabolcsi et al. (2008). They report a set of experiments well designed to prompt a potential facilitation effect of the presence of an NPI on corresponding inferences. They report on both explicit and implicit measures of inference facilitation (mere accuracy in inferential tasks, as well as reading times of phrases that presupposed the conclusion of a downward-entailing type of inference). They report no facilitation effect of the NPI.

In the experiments below we take another look at the question of whether PIs affect monotonicity judgments. Contrary to Szabolcsi et al. (2008), we show that PIs do in fact influence judgments of monotone inferences, just that these effects are (1) only present in cases in which the inferential patterns are less clear to subjects (that is, not in the most basic simple upward or downward-entailing environments); (2) they are not present for all polarity items in all tested configurations; and (3) they push participants more towards rejecting the incorrect inference rather than towards accepting the correct one.

The experimental material, data, the R script used for analysis, as well as the document with the output of all of the models reported in the paper can be found at `https://semanticsarchive.net/Archive/WY4OTMzO`.

3 Experiment 1: PIs affect the perception of monotonicity

Some environments give rise to clear (and correct) judgments of monotonicity: it is quite easy to see that 'John read a novel' entails that 'John read a book'. In such cases, adding a PI may not make the inferences any clearer, or lead people to change their mind in any way about it. In this experiment, we thus looked for an effect on inferences of PIs in contexts in which the inferential patterns are less clear. *Non-monotonic* environments do not support either subset to superset or superset to subset inferences (cf. (7); neither (7a) entails (7b), nor (7b) entails (7a)). However, it has previously been shown that monotonicity judgments of these environments could be more graded, with participants reporting to a non-negligeable extent some monotonicity in one direction or another (see Chemla et al. 2011). Given this level of uncertainty as to whether these environments support upward or downward inferences, we hypothesized that the presence or absence of a PI might then have more room to influence the judgement. Importantly for our purposes, both PPIs and NPIs are known to be acceptable at least to a certain extent in these environments: both (8a) and (8b) can be interpreted as (7a) (there is however some individual variation in terms of NPI acceptability in NM environments, cf. Rothschild 2006, Crnič 2014, Chemla et al. 2011, Denić et al. 2018).

(7) a. Exactly 12 aliens saw birds.
 b. Exactly 12 aliens saw doves.

(8) a. Exactly 12 aliens saw some birds.
 b. Exactly 12 aliens saw any birds.

3.1 Method

3.1.1 Instructions and task

At the beginning of the task, the participants read the following instructions:

(9) *You will see pairs of sentences about aliens, who just spent last week on Earth. Imagine that you hear the first sentence, and indicate whether you would then naturally conclude that the second sentence is true.*

They were then given three examples of such pairs, call them premise-conclusion pairs. In one pair, the conclusion clearly followed from the premise (10), in a second one the conclusion clearly did not follow from the premise (11), and the third case was less clear (12).

(10) 'Each alien received a high score in all human IQ tests.' $\rightarrow$ Aliens are very intelligent.

(11) 'Few aliens visited Paris.' $\rightarrow$ All aliens visited the Eiffel Tower.

(12) 'Pink aliens have scary teeth.' $\rightarrow$ Pink aliens are the most terrifying.

The participants were instructed to record their responses on a continuous scale presented in the form of a bar by filling a portion of it red. They were explained that they could use the flexibility of the red bar to report intermediate judgments, and that they would get used to it naturally. The dependent measure was the percentage of the bar filled in red. This measure will be referred to as the 'rating' given to an inference.

3.1.2 Material

The material was made of pairs of sentences, which were intended to serve as the premise and the conclusion in some inferential judgment task. These pairs of sentences were constructed from the recombination of more atomic building blocks. Crucially, among these pairs there were both valid and invalid upward and downward inferences, with and without polarity items.

The building blocks to create these inferences were as follows. First, we created a list of 11 environments: 3 Upward Entailing (UE) environments (positive, Every, Many), 3 Downward Entailing (DE) environments (negative, No, Few) and 2 Non-Monotonic (NM) environments (Exactly 12, Only 12). Second, we created a list of 12 pairs of (superset, subset) VPs that could host a PI (see <PI> birds, see <PI> doves). We combined these two building blocks, environments and pairs of VPs, to obtain pairs of sentences for our inferential stimuli. Both orders of the pairs were used. Note that only one of the orders provides a valid inference in DE and UE environments, and neither of the orders provides valid inferences in NM environments.

Finally, for each of these pairs of sentences, we created items for which, in the premise, there was (i) no PI (for all environments), (ii) an NPI for DE and NM environments, (iii) a PPI for UE and NM environments. These possibilities correspond to all possibilities that may not be outrageously infelicitous (see discussions about the marginal acceptability of some PIs in NM environments in Rothschild 2006, Crnič 2014, and a quantitative evaluation in Chemla et al. 2011 and Denić et al. 2018).

Overall, we obtained 2 [superset/subset vs subset/superset] x 12 [VPs] x (3 [UE] x 2 [PPI vs no PI] + 3 [DE] x 2 [NPI vs no PI] + 2 [NM] x 3 [NPI vs PI vs no PI]) = 432 inference pairs. One example pair for each environment is provided in (13)-(20).

(13) Condition: UE-positive, superset → subset, (PPI)
  a. The purple alien saw (some) birds.
  b. The purple alien saw doves.

(14) Condition: UE-every, superset → subset, (PPI)
  a. Every alien saw (some) birds.
  b. Every alien saw doves.

(15) Condition: UE-many, superset → subset, (PPI)
  a. Many aliens saw (some) birds.
  b. Many aliens saw doves.

(16) Condition: DE-negative, superset → subset, (NPI)

a. The purple alien didn't see (any) birds.
b. The purple alien didn't see doves.

(17) Condition: DE-no, superset $\rightarrow$ subset, (NPI)
a. No alien saw (any) birds.
b. No alien saw doves.

(18) Condition: DE-few, superset $\rightarrow$ subset, (NPI)
a. Few aliens saw (any) birds.
b. Few aliens saw doves.

(19) Condition: NM-exactly 12, superset $\rightarrow$ subset, (PPI/NPI)
a. Exactly 12 aliens saw (some/any) birds.
b. Exactly 12 aliens saw doves.

(20) Condition: NM-only 12, superset $\rightarrow$ subset, (PPI/NPI)
a. Only 12 aliens saw (some/any) birds.
b. Only 12 aliens saw doves.

These 432 items were distributed in three groups of 144 items each, so that: (a) all 12 VPs would appear in a group, (b) four different VPs were used in items which had a PPI in the premise, four different VPs were used in items which had an NPI in the premise, and four different VPs were used in items which had no PI in the premise, (c) across groups, all 12 VPs would appear with the three types of items (an NPI in the premise, a PPI in the premise, no PI in the premise). Hence, in each group there were 4 [items with different VPs] x 2 [set/subset vs subset/set] x (3 [UE] x 2 [PPI vs no PI] + 3 [DE] x 2 [NPI vs no PI] + 2 [NM] x 3 [NPI vs PI vs no PI]). Participants were administered one of these groups of items, presented each time in a random order.

Apart from the 144 target items, the participants in each group also had to provide responses to the three training items which were administered at the beginning of the task. They were identical to the examples discussed in the instruction, and their purpose was to let participants get used to the setting and to the task.

3.1.3 Participants and exclusion criteria

75 participants were recruited through Amazon Mechanical Turk (38 females). As the result of the following two exclusion criteria, the responses of 66 participants among them were kept for the analysis (32 females). First, the results of one participant were excluded for them reporting not being a native speaker of English. Second, the results from eight more participants were excluded for not judging downward inferences higher in downward than in upward monotonic environments, or for not judging upward inferences higher in upward than in downward monotonic environments. The rationale for this second exclusion criterion is that, as it is likely that these judgments should be straightforward and maximally polarized, these eight participants did not understand the task in

the way we expected (or were responding at random). The same exclusion criteria were applied in all four experiments reported in this paper.

3.2 Results

Responses given in less than 1.4 s (1% of the data) or more than 10s (9% of the data) were removed from the analysis. These numbers were chosen by a visual inspection of the distribution of RTs, with the goal of removing clear outliers. We excluded more responses falling on the slow part of the spectrum, to exclude non-spontaneous responses. This criterion was thus chosen by hand, looking only at the RTs and not the condition and responses they corresponded too. It was then copied without change for the following experiments, which provide replications of these results.

The three training items were answered as expected, with ratings of 93% for the clearly valid inference (10), 9% for the clearly not valid one (11), and 68% for the intermediate one (12).

Fig. 1 represents on the y-axis ratings of inferences with subset in premise and superset in conclusion. These inferences were valid for UE environments. These ratings thus measure the perceived UE-ness of the environment, and we refer to them as UE-ratings. On the x-axis, the ratings correspond to superset to subset inferences, which were valid in DE environments, and are accordingly referred to as DE-ratings. The graph then reports the mean such rating across participants and across the three types of environments (UE, DE, and NM). The graph shows that participants were behaving properly on these broad distinctions: disregarding the effect of PIs for the time being, UE environments ended up in the top left corner of the graph with high UE-ratings ($89.7\%, SD = 13.03$) and low DE-ratings ($29.6\%, SD = 16.2$), DE environments ended up in the bottom right corner of the graph with high DE-ratings ($81.2\%, SD = 15.2$) and low UE-ratings ($31.9\%, SD = 19.9$), and NM environments ended up in the bottom left corner of the graph with low DE and UE-ratings, even if slightly less sharply (respectively, $27.7\%, SD = 17.5$; $44.7\%, SD = 26.3$).

The results are further separated depending on whether the premise contained a PPI, an NPI, or no PI, which is the core manipulation of interest: **does the presence of these items influence UE and DE ratings?** In order to answer this question, we entered ratings in a model by first transforming the ratings so that they would receive a unique directional interpretation: UE-ratings were kept untransformed, but DE-ratings were reversed (x would become $100\% - x$). These transformed measure aligns the ratings across conditions in the following sense: it measures to what extent UE inferences follows, and to what extent DE inferences do not follow. We will refer to these as directional ratings.

Focusing on NM environments, mean participants' directional ratings seem to be, numerically, influenced by the polarity items, and in particular NPIs: while the average aligned ratings without PI and with a PPI are similar ($60\%, SD = 11.7$ and $59.7\%, SD = 12.2$, respectively), the presence of the NPI gave rise to lower directional ratings ($55.3\%, SD = 11.2$). The pertinence of these observations were confirmed by the following (planned) analyses:

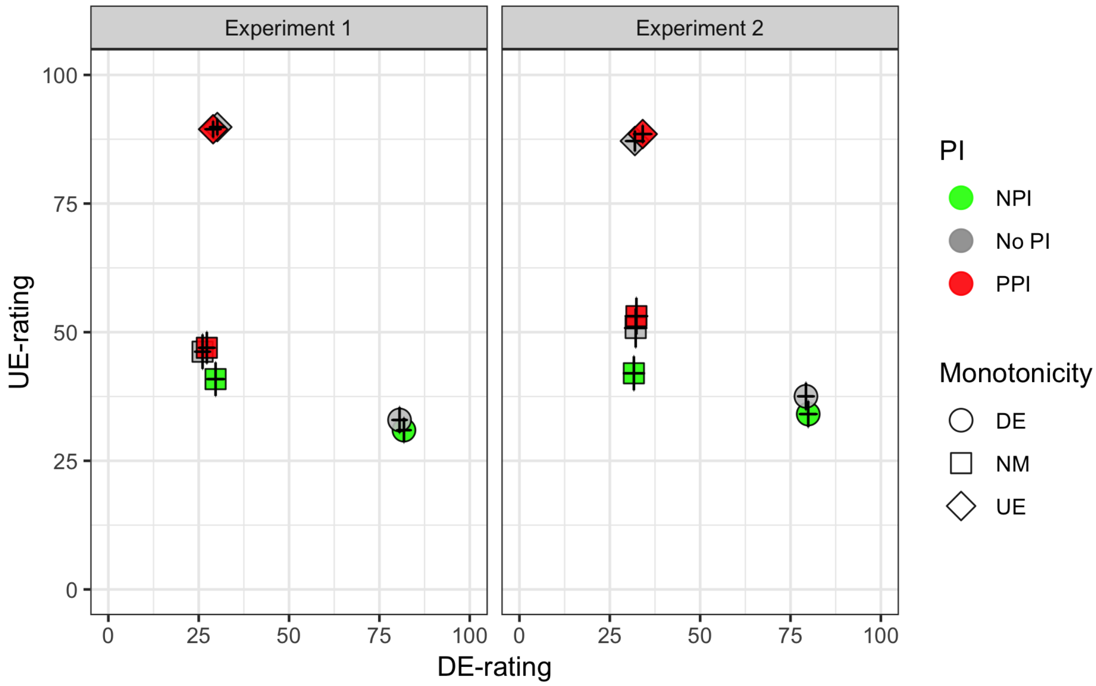

Figure 1: Experiments 1 and 2: Mean participants' rating of the superset to subset inference (DE-rating) and of the subset to superset inference (UE-rating) in DE, UE, and NM environments depending on whether the premise contained a PPI, an NPI, or no PI. Error bars represent standard errors.

The role of NPIs in NM environments The data was subsetted to items with either an NPI or no PI in the premise that is headed by a NM quantifier (Exactly 12, Only 12). A linear mixed model was fitted on this data set's directional ratings with the PI (present or absent), Quantifier (Exactly 12, Only 12) and Inference direction (set/subset vs. subset/set) as fixed effects, and the maximal random effect structure for which convergence was achieved.[1] A comparison of this model with a reduced model without the PI as fixed effect revealed a significant effect of PI on the directional rating ($\chi^2(1) = 18.6, p < .001$). In other words, the presence of an NPI in the premise makes the participants perceive an environment as less upward-entailing as compared to when no PI is in the premise.

The role of PPIs in NM environments The data was subsetted to items with either a PPI or no PI in the premise that is headed by a NM quantifier (Exactly 12, Only 12). A linear mixed model was fitted on this data set's directional ratings with the PI (present or absent), Quantifier (Exactly 12, Only 12) and Inference direction (set/subset vs. subset/set) as fixed effects, and the maximal random effect structure for which convergence was achieved.[2] A comparison of this model with a reduced model with-

[1]The maximal random effects structure for which convergence was achieved included by-participant intercepts and by-participant slopes for Inference direction.

[2]The maximal random effects structure for which convergence was achieved included by-participant and by-item intercept, as well as by participant slopes for Inference direction. Items (here and in the other

out the PI as fixed effect revealed no significant effect of PI on the directional rating ($\chi^2(1) = .05, p = .83$). In other words, the presence of a PPI in the premise did not make the participants perceive an environment as more or less upward-entailing as compared to when no PI is in the premise in this experiment.

3.3 Summary and discussion

Based on the overall pattern of results and non-contentious cases, our inference measure appears to be a relevant measure of the participants' perceived monotonicity of the environment (see also Chemla et al. 2011), according to which DE and UE environments are perceived as such, and NM environments as intermediate. Our crucial result is that in NM environments, the presence of an NPI can lead participants to perceive NM environments as less UE-like. No analogous effect of PPIs was found.

Note that the main observed effect (i.e. the effect of NPIs on inferential judgments in NM environments) cannot be explained away as a regression to the mean (i.e. as more random responses than in the no PI condition leading to judgments overall closer to 50% in the PI condition). The reason why this alternative explanation is not viable is that the main effect of the NPI seems to be in making people perceive a NM as less upward-entailing, pushing them further away from the mean (50%) UE-rating than the baseline (no PI) condition is (cf. Fig. 1).

4 Experiment 2: Replication (with PIs also in conclusions)

The previous result is arguably small in size. Note however that by having environments with and without PIs, in the same experiment, we were likely not to get any result *at all*: participants could have easily figured out that PIs were irrelevant.

Furthermore, in Experiment 1 PIs were present only in the premises, to assess their role as a guide for 'future' inferences, but this creates an asymmetry between premise and conclusion, which may obscure the effect of the PI (for instance, as the PI was present in the premise but not in the conclusion, it might have been quite easy for participants to ignore it.) We thus ran Experiment 2, which was identical to Experiment 1, except that whenever a PI was present in the premise, it was also present in the conclusion.

Roughly put, this Experiment provides us with a replication of the previous result.

4.1 Method

4.1.1 Instructions and task

Instructions and task were identical to those in Experiment 1.

models) corresponded to the VPs used in the sentence.

4.1.2 Material

Material was identical to those used in Experiment 1 except that the conclusion sentences of the inferences contained a PI whenever the premise did.

4.1.3 Participants

72 participants were recruited through Amazon Mechanical Turk (35 females). One participant was excluded from the analysis for reporting not being a native speaker of English and seven more for not showing a difference in perceived monotonicity of upward and downward entailing environments (same criteria as described in section 3.1.3). 64 participants were thus kept for the analysis (28 females).

4.2 Results

To preview, the results of Experiment 2 were qualitatively identical to the results of Experiment 1. As in Experiment 1, responses given in less than 1.4s (6% of the data) or more than 10s (7% of the data) were removed.

Training items were answered as expected: the clearly valid inference received an average rating of 93%, the clearly not valid one 9%, the intermediate one 62%. The right hand side of Fig. 1 summarizes the results from this experiment. As before, it represents on the x-axis mean participants' rating of superset to subset inferences (DE-ratings), and on the y-axis mean participants' rating of subset to superset inferences (UE-ratings), across three types of environments (UE, DE, and NM), depending on whether the premise and conclusion contained a PPI, an NPI, or no PI.

As in Experiment 1, if we first disregard the effect of PIs, the three types of environments UE, DE and NM behave distinctly, as expected. We also observe a similar pattern as before for the role of PIs in NM environments, with inferences without PI and with a PPI receiving similar directional ratings ($59.2\%, SD = 14.9$ and $60.1\%, SD = 14.1$, respectively), and the presence of NPIs leading to lower ratings ($54.7\%, SD = 12.5$).

The role of NPIs in NM environments Linear mixed model comparisons[3] as in Experiment 1 revealed a significant effect of the presence of the NPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 19.04, p < .001$). In other words, we replicate the result from Experiment 1 that the presence of an NPI in the premise makes the participants perceive an environment as less upward-entailing as compared to when no PI is in the premise.

The role of PPIs in NM environments Linear mixed model comparisons[4] as in Experiment 1 revealed no significant effect of the presence of the PPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 1.4, p = .24$). In other words, as it was

[3] Convergence was achieved for the random effects structure with by-participant intercepts and by-participant slopes for Inference direction.

[4] Convergence was achieved for the random effects structure with by-participant intercepts and by-participant slopes for Inference direction.

the case in Experiment 1, the presence of a PPI in the premise did not make the participants perceive an environment as more or less upward-entailing as compared to when no PI is in the premise in Experiment 2 either.

4.3 Summary and discussion

In Experiment 1, we tested whether the presence of NPIs and PPIs as compared to no PIs in the premise has an influence on the monotonicity inferences with conclusions without PIs. Experiment 2 differed from Experiment 1 only in that whenever a PI was present in the premise, it was also present in the conclusion. The results of Experiment 2 confirm those of Experiment 1: the presence of NPIs makes participants consider an environment less upward-entailing than when it contains no PIs, while no similar effect on monotonicity judgments is found for PPIs.

5 Interim discussion

NPIs have been found to influence monotonicity inferences in NM environments: they lead participants to report that these environments are less UE-like than when no PI is present (it seems that the most important part of the effect is in making people perceive NM environments as less UE rather than perceiving them as more DE, as it can be seen in Fig. 1). The effect seems to be more robust in NM environments than in DE environments (cf. Fig. 1), and we believe that this could be well-explained by *ceiling* effects in DE environments (in our experiments, as well as in Szabolcsi et al. 2008). Indeed, for DE environments, participants may report DE judgments as much as they possibly can even without an NPI, leaving little room for an NPI to make the judgments even more extreme.

What should we make of the asymmetry between NPIs and PPIs? As we found no effect of PPIs, we cannot conclude much about the role of monotonicity in PPI licensing. It could be that the monotonicity does not play a role in PPI licensing, as some authors have proposed (Denić 2015, Progovac 2000).

An alternative would be that monotonicity plays a role in PPI licensing too, but that we simply didn't have enough power to capture the effect of PPIs on monotonicity judgments, or that we are again observing ceiling effects. In fact, in both Experiment 1 and Experiment 2 we observe that people judge NM environments as more UE than DE (average endorsement of the UE inferences in NM environments is 44.7% ($SD = 26.3$) in Experiment 1, and 48.6% ($SD = 28.5$) in Experiment 2 , while average endorsement of DE inference in NM environments is 27.7% ($SD = 17.5$) in Experiment 1, and 32.1% ($SD = 22.6$) in Experiment 2.[5] This might have created a ceiling effect for PPIs but not for NPIs: as people perceive NM environments as more UE than DE, NPIs could exercise their ef-

[5]Post-hoc comparison of a linear mixed model comparison fitted on raw responses on NM sentences without PIs with Inference direction as fixed effects with random by-participant intercepts and slopes with a reduced model with just the random effects structure revealed a significant effect of Inference direction both in Experiment 1 ($\chi^2(1) = 37.8, p < .001$) and in Experiment 2 ($\chi^2(1) = 126.35, p < .001$). In Experiment 2, the convergence was achieved with by-participant intercepts only.

fect by making people perceive NM environments as less UE, while there might not be enough room for the same effect of PPIs.

6 Doubly Negative Environments

In Experiments 1 and 2, we have established that there is an influence of polarity items in NM environments. NM environments are characterized by two aspects: first, inferential judgments in these environments are difficult; second both positive and negative polarity items are licensed in these environments, at least to some extent. In the continuation of the paper, we will extend the inquiry to another type of environments with these same properties, albeit possibly for rather different reasons.

The starting point is that NPIs can appear in the so called *doubly-negative* environments that are globally upward-entailing, as in:

(21) Everyone who couldn't see **at all** failed to find the treasure.

Due to the combination of two downward-entailing functions (*n't* is DE and *every* is DE in its restrictor), the NPI *at all* ends up appearing in a global UE environment in (21). This example thus shows that global logical properties cannot (always) be responsible for NPI licensing. Therefore researchers who advocate a monotonicity-based approach to licensing are led to go local, i.e. propose that the system that checks the acceptability of a given PI in a sentence S has access to subconstituents of S, and that PIs are licensed if at least one of the subconstituents they are in has the right monotonicity properties (Gajewski 2005, Homer 2012). For concreteness, in a sentence like (21), this system can single out the VP of the relative clause and compute its monotonicity with respect to the position of the NPI *at all* (monotonicity is a property of functions; to evaluate what we loosely call the monotonicity of a constituent, one has to abstract over a position within this constituent, e.g. the position of the PI): this constituent turns out to be DE w.r.t. this position. As the licensing condition just requires that a PI be in at least one constituent which has the appropriate monotonicity w.r.t. its position, the NPI *at all* is licensed in (21). Note that for the syntactic approach, examples such as (21) are not problematic, as the monotonicity of the environments of PIs is not directly relevant: all that matters is that the NPI be in the right syntactic configuration with at least one appropriate operator.

There is, however, yet another interesting possibility for why NPIs are acceptable in environments like (21). It is possible that monotonicity inferences are so hard in these environments that people wrongly consider them DE to some extent. The NPIs would thus be licensed in these environments because of the subjective (wrong) perception of their monotonicity.

Doubly-negative environments (remember that they are locally DE but globally UE) have two properties in common with NM environments which make them interesting as a further test case for an effect of polarity items on various sorts on inferential judgments: (i) they can host both positive and negative polarity items (see (22)); (ii) even if these environments are plainly upward entailing, we submit that the corresponding inferential judgments are not easy.

Furthermore, this inquiry could also be directly informative about the status of NPI licensing in these environments (are they licensed because the local environment is DE, or because the global environment is wrongly perceived as DE/not UE?). We will discuss this later on, as it will be easier to do so with the results in place.

7 Experiment 3: Doubly Negative is not Positive

This third experiment tested the effect of the PIs on monotonicity inferences in environments with two accumulating downward-entailing operators, such as (22). We will refer to these environments as DN in the continuation, for Doubly Negative environments. As mentioned above, both PPIs and NPIs are known to be licit in these environments (which means, for PPIs like *some*, that they can be interpreted with narrow scope under both operators; for instance, (22a) can be interpreted as (22b)):

(22) a. Every alien who did not see some doves is hairy.
b. Every alien who did not see any doves is hairy.

7.1 Method

7.1.1 Instructions and task

Instructions and task were identical to those in Experiment 1 and 2.

7.1.2 Material

The stimuli were identical to those used in Experiment 1, except for the addition of two DN environments. These were presented with a PPI, an NPI, or no PI in the premise (and no PI in the conclusion). We thus obtained 2 [superset/subset vs subset/superset] x 12 [VPs] x (3 [UE] x 2 [PPI vs no PI] + 3 [DE] x 2 [NPI vs no PI] + 2 [NM] x 3 [NPI vs PI vs no PI] + 2 [DN] x 3 [NPI vs PI vs no PI]) = 576 inference pairs. An example of premise-conclusion pair for each of the two DN environments is in (23) and (24). These 576 items were split into three groups with 192 items, which satisfied the same conditions as the groups in Experiment 1. Participants were randomly administered to one of the three groups.

(23) Condition: DN-Every-not, subset $\rightarrow$ superset, (PPI/NPI)
a. Every alien who did not see (some/any) doves is hairy.
b. Every alien who did not see birds is hairy.

(24) Condition: DN-No-without, subset $\rightarrow$ superset, (PPI/NPI)
a. No alien spent a year without seeing (some/any) doves.
b. No alien spent a year without seeing birds.

7.1.3 Participants

112 participants were recruited through Amazon Mechanical Turk (69 females). Seven participants were excluded from the analysis for reporting not being a native speaker of English and 13 more for not showing much difference in perceived monotonicity of upward and downward entailing environments (same exclusion criteria as in Experiments 1 and 2). 92 participants were thus kept for the analysis (53 females).

7.2 Results

To preview, in Experiment 3 we mostly replicate the result from Experiments 1 and 2 for the NM environments. As for DN environments, no effect of the presence of NPIs on the global monotonicity inference was found, but interestingly, an effect of the PPI was found.

As in Experiments 1 and 2, responses given in less than 1.4s (4% of the data) or more than 10s (13% of the data) were removed from the analysis. Training items were answered as expected: the clearly valid inference received an average rating of 92%, the clearly not valid one 8.3%, the intermediate one 67%.

Fig. 2 represents on the x-axis mean participants' rating of superset to subset inference (DE-rating), and on the y-axis mean participants' rating of subset to superset inference (UE-rating), across four types of environments (UE, DE, NM, and DN), depending on whether the premise contained a PPI, an NPI, or no PI.

Disregarding whether and which PI was present in the premise, the four types of environments (UE, DE, NM and DN) are well-separated and they show up where they could have been expected. DN environments are quite interesting in this respect: as a reminder, DN environments are in fact plain UE environments. Nonetheless, they seem to behave in a more intermediate fashion, and they are much closer to NM than to UE environments.

Looking first at the replication of the results from Experiments 1 and 2 in NM environments, mean participants' directional ratings were (i) $55.9\%(SD = 13.4)$ when NPI is in the premise, (ii) $57.9\%(SD = 14)$ when PPI is in the premise, and (iii) $57.1\%(SD = 13.3)$ when no PI is in the premise.

The role of NPIs in NM environments Linear mixed model comparisons[6] as in Experiment 1 revealed a borderline effect of the presence of the NPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 2.95, p = .086$) in NM environments. In other words, even though Experiment 3 is like Experiments 1 and 2 in that numerically the NPI in the premise makes the participants perceive NM environments as less upward-entailing as compared to when no PI is in the premise, this effect doesn't reach significance in Experiment 3, unlike in Experiments 1 and 2.

[6]Convergence was achieved for the random effects structure which included by-participant intercepts only.

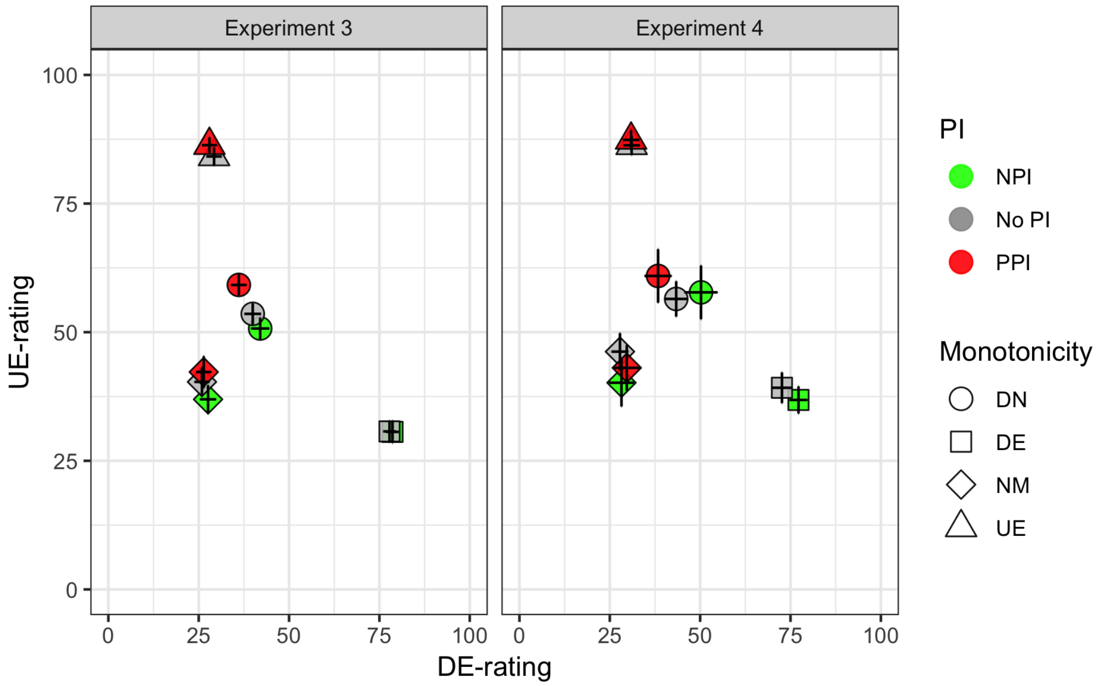

Figure 2: Experiment 3 and 4: Mean participants' rating of the superset to subset inference (DE-rating) and of the subset to superset inference (UE-rating) in DE, UE, NM, and DN environments depending on whether the premise contained a PPI, an NPI, or no PI. Error bars represent standard errors.

The role of PPIs in NM environments Linear mixed model comparisons[7] as in Experiments 1 and 2 revealed no significant effect of the presence of the PPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 1.9, p = .17$) in NM environments. In other words, as it was the case in Experiments 1 and 2, the presence of a PPI in the premise did not make the participants perceive non-monotonic environment as more or less upward-entailing as compared to when no PI is in the premise.

Moving to DN environments, mean participants' directional ratings were (i) $53.7\%(SD = 16.8)$ when NPI is in the premise, (ii) $61.8\%(SD = 14.4)$ when PPI is in the premise in, and (iii) $56.9\%(SD = 14.7)$ when no PI is in the premise.

The role of NPIs in DN environments Linear mixed model comparisons as for the NM environments was done for DN environments[8] and it revealed no significant effect of the presence of the NPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 1.96, p = .16$). In other words, we find no effect of NPI on monotonicity inferences in DN environments.

[7]Convergence was achieved for the random effects structure which included by-participant intercepts and by-participant slopes for Inference direction.

[8]Convergence was achieved for the random effects structure which included by-participant and by-item intercepts and by-participant slopes for Quantifier and Inference direction.

The role of PPIs in DN environments Linear mixed model comparisons[9] revealed a significant effect of the presence of the PPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 11.8, p < .001$). In other words, even though the presence of PPIs had no effect on monotonicity inferences in NM environments in Experiments 1-3, it did influence monotonicity inferences in DN environments.

7.3 Summary and discussion

In Experiment 3, in addition to replicating the results from Experiments 1 and 2 (although in one case, only numerically), we tested whether the presence of NPIs and PPIs as compared to no PIs in DN environment has an influence on global monotonicity inferences. We found no significant effect of NPIs, but an effect of PPIs on the global monotonicity inferences was found in DN environments (which are upward entailing). Why do we see the effect PPIs on inferential judgments in DN but not in NM environments?

This could be related to the fact that the baseline inferential judgments may leave more room for PPI effects in DN environments than in NM environments. Let us explain. While neither of the two environments are DE, DN environments are judged on average as more DE than NM environments (average DE-rating in DN is 40.1%, with $SD = 23.7$, and average DE-rating in NM environments is 28.5%, with $SD = 20.4$). This means that there is more room for improvement in terms of DE-rating in DN environments as opposed to NM environments. Accordingly, a PPI effect pushing judgments away from this mistake may be easier to observe for DN environments.

However, this result must be interpreted with care, as there is an alternative explanations for why we see the effect of PPIs in DN but not in NM environments. This alternative explanation comes from the fact that PPIs like *some* can take an exceptional wide scope. For instance, (25a) has an interpretation according to which *some* takes the widest scope in the sentence (this interpretation is paraphrased in (26)). Note that under this interpretation, the subset to superset inference from (25a) to (25b) still follows.

(25) a. Every alien who did not see some doves is hairy.
b. Every alien who did not see birds is hairy.

(26) Some doves are such that every alien who didn't see them is hairy.

This means that, if for whichever reason the wide scope of *some* is easier to obtain in DN environments than in NM environments, and if the wide scope interpretation of *some* makes it easier to see that the inference from (25a) to (25b) follow, one could explain the effect of PPI in DN environments without relating PPI licensing to monotonicity inferences.

Let us now discuss another puzzling aspect of the results from Experiment 3, namely, that we don't see an effect of NPIs in DN environments. Why would this be?

[9] Convergence was achieved for the random effects structure which included by-participant intercepts and by-participant slopes for Quantifier and Inference Direction.

The first, uninteresting, possibility, is that we simply didn't have enough power to capture it. As DN environments are arguably the most complex among the environments tested in the experiments, they might have lead to a larger within-participant variation (this seems to clearly be the case in Experiment 4, even though less so in Experiment 3), requiring more power to capture the effect if it was there.

There is, however, yet another possibility, compatible with theories according to which an NPI is licensed if there is at least one constituent which has the right downward monotonicity properties. If these theories are on the right track, the NPI in DN environments can in principle have conflicting influences, depending on the constituent in which the NPI is thought to be licensed: (i) it might influence the perception of the local environment as more DE, thus leading to the overall better perception of the global environment as more UE than in the no PI condition, or (ii) it might influence the perception of the global environment as more DE than in the no PI condition. If both are happening to more or less the same extent, it is in fact not surprising not to find a significant effect of the NPI in DN environments.

8 Experiment 4: Doubly Negatives, a replication (with a different arrangement of the items)

8.1 Method

8.1.1 Instructions and task

Instructions and task were identical to those in Experiments 1, 2 and 3.

8.1.2 Material

Materials were the same as those used in Experiment 3, with two differences. First, the total number of items was reduced to 480 by reducing the number of different VPs from 12 to 10. Second, the participants were split into four groups (instead of three) in such a way that each participant sees a given environment either with an NPI, or with a PPI, or without a PI. Because of this, there were 2 [superset/subset vs subset/superset] x 10 [VPs] x (3 [UE] + 3 [DE] + 2 [NM] + 2 [DN]) = 200 items per group.

8.1.3 Participants

81 participants were recruited through Amazon Mechanical Turk (43 females). Four participants were excluded from the analysis for reporting not being a native speaker of English and six more for not showing much difference in perceived monotonicity of upward and downward entailing environments (same exclusion criteria as in Experiments 1-3). 71 participants were thus kept for the analysis (36 females).

8.2 Results

As in Experiments 1-3, we removed responses given in less than 1.4s (8% of the data) or more than 10s (10% of the data). Training items were answered as expected: the clearly valid inference received an average rating of 89.5%, the clearly not valid one 8.2%, the intermediate one 60.7%.

Fig. 2 represents on the x-axis mean participants' rating of superset to subset inference (DE-rating), and on the y-axis mean participants' rating of subset to superset inference (UE-rating), across four types of environments (UE, DE, NM, and DN), depending on whether the premise contained a PPI, an NPI, or no PI.

Disregarding for the time being the role of the PIs on inferential judgments, the four environments (DE, UE, NM, DN) are well separated: these results seem to be qualitatively identical to the results in Experiment 3.

Looking again first at the effect of PIs in NM environments, mean participants' directional ratings are (i) $56.3\%(SD = 11.6)$ when NPI is in the premise, (ii) $56.4\%(SD = 11.1)$ when PPI is in the premise in, and (iii) $59.3\%(SD = 15.3)$ when no PI is in the premise.

The role of NPIs in NM environments Linear mixed model comparisons[10] as in Experiments 1-3 revealed a significant effect of the presence of the NPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 7.8, p < .01$) in NM environments. In other words, we replicate the results from Experiments 1 and 2 that the presence of an NPI in the premise makes the participants perceive NM environments as less upward-entailing as compared to when no PI is in the premise.

Furthermore, with the results of the four experiments in place, we conducted a post-hoc meta-analysis over the four Experiments. Linear mixed model comparisons[11] as above revealed a significant effect of the presence of the NPI (as compared to no PI at all) on monotonicity inferences in NM environments when pooling the results of all experiments ($\chi^2(1) = 38.4, p < .001$).

The role of PPIs in NM environments Linear mixed model comparisons[12] as in Experiments 1-3 revealed no significant effect of the presence of the PPI (as compared to no PI at all) on monotonicity inferences ($\chi^2(1) = 2.6, p = .10$) in NM environments. In other words, as it was the case in Experiments 1-3, the presence of a PPI in the premise did not make the participants perceive NM environments as more or less upward-entailing as compared to when no PI is in the premise.

Again, with the results of the four experiments in place, we conducted a post-hoc meta-analysis over the four experiments. Linear mixed model comparisons[13] re-

[10] Convergence was achieved for the random effects structure which included by-participant and by-item intercepts as well as by-participant slopes for Inference direction.

[11] Convergence was achieved for the random effects structure which included by-participant and by-item intercepts as well as by-participant slopes for Inference direction and for PI.

[12] The maximal random effects structure for which convergence was achieved had by-participant intercepts as well as by-participant slopes for Inference direction.

[13] Convergence was achieved for the random effects structure which included by-participant intercepts as well as by-participant slopes for Inference direction.

vealed no significant effect of the presence of the PPI (as compared to no PI at all) on monotonicity inferences in NM environments across four experiments ($\chi^2(1) = 0.8, p = .38$).

In DN environments, mean participants' directional ratings were (i) $52.6\% (SD = 25.2)$ when NPI is in the premise, (ii) $61.6\% (SD = 17.4)$ when PPI is in the premise in, and (iii) $56.3\% (SD = 22.4)$ when no PI is in the premise.

The role of NPIs in DN environments Linear mixed model comparisons as for the NM environments was done for DN environments. This time, the result was ambiguous, as convergence was achieved for two equally complex random effects structures. According to the first one[14], the presence of the NPI (as compared to no PI at all) has a significant effect on monotonicity inferences ($\chi^2(1) = 27.9, p < .001$). According to the second one[15], it does not ($\chi^2(1) = 2.62, p = .11$).
We conducted a post-hoc meta-analysis to evaluate this effect pooling Experiments 3 and 4, in which these environments were tested. Linear mixed model comparisons[16] revealed a significant effect of the presence of the NPI (as compared to no PI at all) on monotonicity inferences in DN environments ($\chi^2(1) = 7.2, p < .01$).

The role of PPIs in DN environments Linear mixed model comparisons lead to an ambiguous result, as convergence was achieved for two equally complex random effects structures. According to the first one[17], the presence of the PPI (as compared to no PI at all) has a significant effect on monotonicity inferences ($\chi^2(1) = 22.3, p < .001$). According to the second one[18], it does not ($\chi^2(1) = 2.1, p = .15$).
We conducted a post-hoc meta-analysis to evaluate this effect pooling Experiments 3 and 4, in which these environments were tested. Linear mixed model comparisons[19] revealed a significant effect of the presence of the PPI (as compared to no PI at all) on monotonicity inferences in DN environments ($\chi^2(1) = 10.7, p < .01$).

8.3 Summary and discussion

In Experiment 4, we replicate the effect of the presence of the NPI on monotonicity inferences in NM environments. This effect is further supported by the meta-analysis performed on the results of the four experiments reported in this paper. As in the previous

[14] In the first model, convergence was achieved with the random effects structure which included by-participant intercepts and by-participant slopes for Inference direction.

[15] In the second model, convergence was achieved with the random effects structure which included by-participant intercepts and by-participant slopes for Quantifier.

[16] Convergence was achieved for the random effects structure which included by-participant intercepts as well as by-participant slopes for Inference direction and PI.

[17] In the first model, convergence was achieved with the random effects structure which included by-participant intercepts and by-participant slopes for Inference direction.

[18] In the second model, convergence was achieved with the random effects structure which included by-participant intercepts and by-participant slopes for PI.

[19] Convergence was achieved for the random effects structure which included by-participant and by-item intercepts as well as by-participant slopes for Inference direction and PI.

experiments, in Experiment 4 we find no effect of PPI on monotonicity inferences in NM environments (and a meta-analysis does not say differently).

As for DN environments, the results of Experiment 4 provide some evidence for the effect of both NPIs and PPIs on inferences, which evidence is further confirmed by the joint analysis of Experiment 3 and 4, suggesting that the NPI effect was missed in Experiment 3 for power reasons. However, it is important to remember that this effect currently enjoys lower replicability than other effects reported across these experiments.

9 Conclusion and discussion

In the four experiments reported in this paper, it was found that polarity items affect reasoning judgments. These PI manipulations are subtle on the surface, and give rise to accordingly small and subtle effects on upward and downward monotonicity inferences. These effects were not found in previous investigations with simpler cases (plain DE and UE environments, see Szabolcsi et al. 2008), and likewise, we found that the effects are smaller or absent in those environments. However, in more complex cases where inferential judgments are more difficult, in particular in NM environments, polarity items have been found to influence inferential (monotonicity) judgments across multiple experiments. These results thus reveal the influence on reasoning tasks of subtle and apparently minor choices of closed class words.

Interestingly, the current results also document the fact that monotonicity inferences may be rather difficult to assess in inferential judgments tasks (Geurts and van Der Slik 2005). Whether this is simply a consequence of performing the experimental task in question is open for discussion, but surely it raises challenges for the question of how people derive and understand the truth-conditions of sentences for efficient communication, noting that as soon as one understands the truth-conditional meaning of a sentence, one should be able to see what is entailed by it. We note here that polarity items can help filter out some misunderstandings, if they can help out signaling monotonicity properties.

A final, more technical issue one may raise is how our results bear on the question of why NPIs are licensed in DN environments. There is, in fact, a surprising result which might speak to this question. We found that the environments in which NPIs are thought to be licensed are not all perceived as downward entailing, but all of them are perceived as 'non upward entailing'. This is perhaps unsurprising for DE and NM environments, but it is remarkable for DN environments, which are in fact upward entailing. Furthermore, some results suggest that the presence of an NPI makes one perceive a DN environment as more downward-entailing (which, again, it is not). This opens the possibility that the acceptability of NPIs in such environments is not (only) due to a syntactic relationship with a negative licensor or to the downward monotonicity of the local environment, but rather (at least in part) to the perception that the environment is, at a global level, not upward entailing. This is a significant departure from current approaches (although see Chemla et al. 2011). This perspective emerges here from the systematic collection of inferential and acceptability judgments, and it could naturally be put to further tests (although testing more environments in this way is more costly than through traditional introspective judgments). This theoretical option may also illustrate a particular type of cognitive approach to linguistic generalizations in general, in which subjective and potentially 'fallacious' judgments have their say in grammar and grammatical theorizing.

References

Emmanuel Chemla, Vincent Homer, and Daniel Rothschild. Modularity and intuitions in formal semantics: The case of polarity items. *Linguistics and Philosophy*, 34(6):537–570, 2011.

Gennaro Chierchia. *Logic in grammar: Polarity, free choice, and intervention*. OUP Oxford, 2013.

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

Milica Denić. Minimizing Scope Ambiguity Hypothesis. In *ESSLLI 2015 Student Session*, pages 54–65, 2015.

Milica Denić, Emmanuel Chemla, and Lyn Tieu. Intervention effects in NPI licensing: a quantitative assessment of the scalar implicature explanation. *Glossa: a journal of general linguistics*, 3(1):49, 2018.

Heiner Drenhaus, Peter beim Graben, Douglas Saddy, and Stefan Frisch. Diagnosis and repair of negative polarity constructions in the light of symbolic resonance analysis. *Brain and language*, 96(3):255–268, 2006.

Heiner Drenhaus, Douglas Saddy, and Stefan Frisch. Processing negative polarity items: When negation comes through the backdoor. In S. Kepser and M. Reis, editors, *Linguistic evidence: Empirical, theoretical, and computational perspectives*, pages 145–165. Berlin: Mouton de Gruyter, 2005.

Drenhaus, Heiner, Blaszczak, Joanna, and Schütte, Julianne. Some psycholinguistic comments on NPI licensing. In *Proceedings of Sinn und Bedeutung*, volume 11, pages 180–193, 2007.

Gilles Fauconnier. Polarity and the Scale Principle. *Chicago Linguistics Society*, 11:188–199, 1975.

Jon Robert Gajewski. *Neg-raising: Polarity and presupposition*. PhD thesis, Massachusetts Institute of Technology, 2005.

Bart Geurts and Frans van Der Slik. Monotonicity and processing load. *Journal of semantics*, 22(1):97–117, 2005.

Elena Guerzoni. Intervention effects on NPIs and feature movement: Towards a unified account of intervention. *Natural Language Semantics*, 14(4):359–398, 2006.

Elena Herburger and Simon Mauck. A new look at Ladusaw?s puzzle. In Hedde Zeijlstra and Jan-Philipp Soehn, editors, *Workshop on Negation and Polarity*, pages 78–84, 2007.

Vincent Homer. Domains of polarity items. 2012.

Nirit Kadmon and Fred Landman. Any. *Linguistics and philosophy*, 16(4):353–422, 1993.

Manfred Krifka. Some remarks on polarity items. *Semantic universals and universal semantics*, pages 150–189, 1991.

Manfred Krifka. The semantics and pragmatics of polarity items. *Linguistic analysis*, 25 (3-4):209–257, 1995.

William Ladusaw. *Polarity Sensitivity as Inherent Scope Relations*. PhD thesis, University of Texas Austin, 1979.

Utpal Lahiri. Focus and Negative Polarity in Hindi. *Natural Language Semantics*, 6:57–123, 1998.

Hanna Muller and Colin Phillips. *Negative polarity illusions*, volume Oxford Handbook of Negation. Oxford University Press, 2018.

Dan Parker and Colin Phillips. Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, 157:321–339, 2016.

Ljiljana Progovac. Negative and positive feature checking and the distribution of polarity items. *Negation in Slavic*, pages 88–114, 2000.

Daniel Rothschild. Non-monotonic NPI-licensing, definite descriptions, and grammaticalized implicatures. In *Semantics and Linguistic Theory*, volume 16, pages 228–240, 2006.

Douglas Saddy, Heiner Drenhaus, and Stefan Frisch. Processing polarity items: Contrastive licensing costs. *Brain and language*, 90(1-3):495–502, 2004.

Jenny Shao and Helen Neville. Analyzing semantic processing using event-related potentials. *Newsletter for the Center for Research in Language*, 11(5):3–20, 1998.

Karsten Steinhauer, John E Drury, Paul Portner, Matthew Walenski, and Michael T Ullman. Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia*, 48(6):1525–1542, 2010.

Anna Szabolcsi. Positive polarity–negative polarity. *Natural Language & Linguistic Theory*, 22(2):409–452, 2004.

Anna Szabolcsi, Lewis Bott, and Brian McElree. The Effect of Negative Polarity Items on Inference Verification. *Journal of Semantics*, 25:411–450, 2008.

Shravan Vasishth, Sven Brüssow, Richard L Lewis, and Heiner Drenhaus. Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32(4): 685–712, 2008.

Ming Xiang, Brian Dillon, and Colin Phillips. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, 108(1):40–55, 2009.

Ming Xiang, Julian Grove, and Anastasia Giannakidou. Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*, 4:708, 2013.

Aydogan Yanilmaz and John E Drury. Prospective NPI licensing and intrusion in Turkish. *Language, Cognition and Neuroscience*, 33(1):111–138, 2018.

Anna Yurchenko, Dirk-Bart Den Ouden, Jack Hoeksema, Olga Dragoy, John CJ Hoeks, and Laurie A Stowe. Processing polarity: ERP evidence for differences between positive and negative polarity. *Neuropsychologia*, 51(1):132–141, 2013.