The selectivity and interpretation of NPI illusions

Abstract

This study investigates the illusory licensing of Negative Polarity Items (NPIs), a subtype of linguistic illusion whose behavior is informative for understanding real-time interpretation. We present the results from six experiments, including online/offline judgments and interpretation tasks, in an attempt to identify the processing error that underlies this brief deviation from the mental grammar. The results indicate that illusions are robust for intrusive quantificational negation such as "no authors" but reduced or absent for other forms of intrusive negation such as "didn't". We additionally find that although NPI illusion sentences are often interpreted as expressing a globally negative proposition, these interpretive errors are a consequence of the illusion, not a likely cause of it. These findings pose problems for current accounts which do not predict the error profile nor the interpretation patterns that we observe, including accounts that appeal to the erroneous retrieval of a non c-commanding negative item in memory (Vasishth et al. 2008), a pragmatic rescuing operation (Xiang et al. 2009) or the parser's failure to assign an appropriate scope for negative quantifiers (Orth et al. 2020a). Alternatively, we argue for an account of NPI illusions that focuses on how sentences are interpreted in relation to scalar alternatives, invoking a style of grammatical explanation in the spirit of Fauconnier (1975a, 1975b), which emphasizes the critical role of the relation between the NPI and the context in which it appears.

Keywords: polarity, negation, NPI illusions, scalar inferences, acceptability judgments, real-time interpretation processes.

1 Introduction

The licensing of Negative Polarity Items (henceforth NPIs) like *any* and *ever* has long been a valuable model system in semantic theory. In this project we investigate the illusory licensing of NPIs, a subtype of linguistic illusion whose behavior can be equally valuable for understanding real-time interpretation. The NPI illusion is the fleeting perception of acceptability for ungrammatical sentences with an unlicensed NPI like (1a), in contrast with similar sentences like (1b), whose deviance is immediately detected.

(1) a. *The bills that no senators voted for have ever become law.
b. *The bills that the senators voted for have ever become law.

The illusion has been found to be robust across measurements and languages; yet, most existing investigations have studied NPI illusions using quantificational forms of negation, predominantly *no,* but also other quantificational elements such as *few* or *only*. Here we present the results from six experiments in an attempt to identify the processing error that underlies this brief deviation from the grammar. Specifically, these experiments investigate the illusory potential of a different type of licensors – namely, verbal negation in the form of *not* and *-n't.* The results obtained pose problems for current accounts and suggest an alternative explanation that builds on two independent sets of findings: a) linguistic approaches to the grammar of NPIs that emphasize the importance of scalar inferences, and b) processing accounts of the interpretation of negated statements. Given that we are interested in the online implementation of grammatical knowledge of NPIs, we begin with a brief review of the main linguistic hypotheses about this component of the grammar. We then turn to prior psycholinguistic hypotheses which attempt to explain the NPI illusion, and their relation to the hypothesized grammatical knowledge.

1.1 The grammar of NPIs

The vast majority of words in the lexical inventories of natural languages can occur in both positive and negative contexts. However, languages also have lexical items whose distribution is restricted by the polarity of the context in which they appear, known as *polarity items*. The most salient subtypes of these are positive polarity items (PPIs) and negative polarity items (NPIs). The class of English NPIs includes adverbs like *ever*, *anymore*, *yet* or *in years*, the determiner *any*, noun phrases such as *a red cent* or *a thin dime* and verb phrase idioms like *lift a finger*, *have a hope in hell* and so forth. Roughly (we will refine this definition later) these elements must occur in the scope of a negative element, and thus, are unacceptable in positive contexts (cf. (2a/b)). Note that mere linear precedence of a negative element with respect to the NPI is not enough: the NPI must be under the scope of the negative element (often understood as syntactic c-command) (Laka, 1994). For example, in (2c), the negative element *wasn't,* inside the relative clause, is structurally irrelevant to the NPI, resulting in ungrammaticality.

(2) a. **No** student has **ever** complained about the coursework.
b. *The student has **ever** complained about the coursework.
c. *The student who **wasn't** in class has **ever** complained about the coursework.

A prominent focus of NPI research is what Ladusaw (1996) called *the licensing question*. That is, to provide a unified analysis of the collection of elements and contexts that license NPIs. While the canonical NPI licensor is explicit negation, the class of NPI-licensing environments is in fact much broader. NPIs can be found in questions, comparative structures, the scope of adversative predicates, the antecedent of a conditional, and many other contexts. Identifying the property that these contexts have in common – and accounting for variation across languages and across NPIs within a language – has been a primary focus of research on the grammar of NPIs.

One group of proposals emphasizes a key role for explicit negation in NPI licensing; in this framework, licensing is a syntactic relation between the NPI and an overt, c-commanding negative feature. In order to accommodate the acceptability of NPIs in other contexts, a secondary, indirect mechanism is postulated. This allows for an NPI-containing sentence without explicit negation to be accepted in virtue of its close relation to a sentence which does contain negation, as in (3). The details of that relation vary across proposals. For instance, Baker (1970) emphasizes the importance of logical entailments, while Linebarger (1987) argues that the key relation is that of implicature, and for Giannakidou's (2006) "rescuing" operation, it is any proposition "made available" by the global context.

(3) I doubt that you will **ever** pass the exam = I **don't** think that you will **ever** pass the exam.

A contrasting idea simultaneously emerged in the works of Fauconnier (1975a, 1975b), which treats the distribution of NPIs as only one case study in the broader phenomenon of semantic and pragmatic polarity. The key idea lies in the observation that some sentential contexts are associated with scales in virtue of our world knowledge – if *Ann did not do x to help* and *y* is more effort than *x*, we might typically infer that *Ann did not do y to help* (importantly for Fauconnier, this is not an entailment, merely an implicature). Thus, if we take *lift a finger* to indicate the minimum possible effort, then *Ann did not lift a finger to help* will, in virtue of the associated scale, imply that *Ann did not do x to help* for all other values of *x*. Flipping the polarity of the sentence has the effect of reversing the scale – if *Ann did x to help* and *y* is more effort than *x*, we cannot infer *Ann did y to help*. Thus, an NPI like *lift a finger*, which indicates the minimum effort, will not carry any implications about the rest of the scale in a scale-reversed (positive) context, and so the NPI loses its idiomatic, quantificational reading in the sentence *Ann lifted a finger to help*. Note that while this account does not exactly provide an answer to the licensing question as stated above, it does predict that any context in which a polarity item is acceptable will fail to license the same polarity item if its polarity is reversed.

This key intuition has been pursued in two different approaches to NPI licensing – those that emphasize the meaning of the licensor and those that emphasize the meaning of the NPI. Ladusaw (1980, 1996) and subsequent work in the first category formalized NPI licensing as scope by a downward-entailing operator. "Downward entailment" is simply entailment from more general to more specific statements, such as the inference from (4a) to (4b). Note that this entailment does not hold if the negative operator *not* is removed from both sentences.

(4) a. The students have not complained about the coursework.
b. The students have not complained loudly about the coursework.

Though there is clearly a resemblance between Fauconnier's scales and Ladusaw's downward entailment, the theories are not isomorphic. They differ in their focus on pragmatic versus truth conditional aspects of meaning, as well as the locus of licensing. For Fauconnier it is the scalar alternatives themselves that allow the NPI to be successfully interpreted in context, whereas NPIs in Ladusaw's framework are licensed by a scope relation to a licensor. Thus, an online NPI-licensing mechanism that is faithful to the grammar will require the rapid computation of different types of information under these two hypothesized grammars. The question of what is computed online will be central to hypotheses about what drives illusions.

A separate body of research, also building on Fauconnier's observation about pragmatic scales, has highlighted aspects of the NPI's meaningful contribution to the sentence. Kadmon & Landman's (1993) influential analysis of the NPI *any* proposes that the function of NPIs is to strengthen the claim expressed by the sentence in which they occur. They specifically argue that *any* is used to widen the domain of quantification, and that for the NPI to be pragmatically felicitous, this widening effect must strengthen the statement. Following this current, some authors implemented the notions of widening and strengthening by focusing on the relevance of subdomain alternatives and scalar inferences in the occurrence of NPIs (e.g. Krifka 1995; Israel 1997, 2011; Chierchia 2006). The central idea of these approaches, following Fauconnier's initial insight, is that the meanings of NPI-containing sentences correspond to extreme values along a scale of ordered alternatives that can be contextually inferred.

From this brief presentation of the main theoretical approaches to the grammar of NPI licensing, it is clear that negative polarity phenomena lie at the interface of syntactic, semantic and pragmatic mechanisms. While we should be cautious about selecting among competing grammatical theories on the basis of sentence processing data, we note that some formulations of the language user's knowledge lend themselves more to some processing mechanisms. For example, grammatical hypotheses that define licensing as a syntactic relation between an NPI and the features of a prior licensor might be straightforwardly implemented as a memory retrieval operation of that prior negative word. In contrast, hypotheses that treat licensing as an operation that relates the NPI-containing sentence to its alternatives would instead require a process by which those alternatives are activated and compared. Against this backdrop, we now turn to existing research on the illusory processing of NPIs.

1.2 Accounts of NPI illusions

While there is much diversity in extant accounts of the mechanism underlying NPI licensing, the grammatical accounts of NPI licensing explored above all predict that (5a) is grammatical and (5b) and (5c) are ungrammatical in virtue of the lack of an appropriate licensor for the NPI *ever* (examples from Parker and Phillips 2016). Informal acceptability judgments typically align with this prediction – native speakers accept (5a) and reject (5b) and (5c). Critically, however, (5b) and (5c) are not alike in online measures such as speeded acceptability. Comprehenders sometimes fail to detect the ungrammaticality of sentences like (3b), leading to a grammatical illusion.

(5) a. **No** authors [that the critics recommended] have **ever** received …
b. *The authors [that **no** critics recommended] have **ever** received …
c. *The authors [that the critics recommended] have **ever** received …
… acknowledgement for a best-selling novel.

The existence of illusion effects in the processing of unlicensed NPIs is an empirically robust phenomenon, both across languages and measurements. It has been replicated using methods such as speeded acceptability judgments (*German*: Drenhaus et al. 2005; *English*: Xiang et al. 2006; Parker & Phillips 2016; de-Dios-Flores et al. 2017; Hildebrandt & Husband 2019; Muller et al. 2019; Orth et al. 2020a; *Korean*: Yun et al. 2018), self-paced reading (*English*: Parker & Phillips 2011, 2016; Xiang et al. 2013; Ng & Husband 2017; *Turkish*: Yanilmaz & Drury 2018a), eye-tracking (*German*: Vasishth et al. 2008; *English*: Orth et al. 2020b) and event-related potentials (*German*: Drenhaus et al. 2005; *English*: Xiang et al. 2009; *Turkish*: Yanilmaz & Drury 2018b; Lee et al. 2018; *Korean*: Lee et al. 2018). Yet, the processes that lead to the facilitation of ungrammatical illusion sentences are still not well understood. We take the key question to be why a comprehension system equipped with a grammar of NPI licensing appears to not respect these grammatical constraints in initial stages of processing. This framing question forces us to consider not only the assumed grammatical knowledge but also the possible mechanisms for implementing grammatical knowledge in a rapid incremental comprehension system.

There are two existing influential approaches to the NPI illusion: one that conceptualizes NPI licensing as fundamentally a memory retrieval operation and places blame for illusions on properties of the memory architecture (Vasishth et al. 2008), and one that highlights pragmatic inferences such as a “rescuing” operation for NPIs not licensed by explicit negation and places blame for illusions on overzealousness in this system (Xiang et al. 2009). We will refer to these as the *memory-based hypothesis* and the *pragmatic rescuing hypothesis*, respectively.

Under the memory-based hypothesis, memory retrieval operations are executed via parallel cue-based activation of content-addressable items (Lewis & Vasishth 2005). The licensing of an NPI in real time is re-framed as a problem of retrieving a licensor from the memory store of the preceding sentence fragment. In this model, the successful retrieval of an item in memory is the result of the item’s level of activation and the item’s feature-by-feature match to the retrieval cues. For NPI licensing, [+negation] and [+c-command] have been suggested as candidate retrieval cues, but other cue combinations are possible. For example, [+downward entailing] and [+scope] would be appropriate features in a processing theory that hews more closely to Ladusaw’s (1979) downward entailing hypothesis. There are notable challenges in implementing relational notions like c-command and scope as retrieval cues, but these do not impact the current discussion. The key component of the memory hypothesis for explaining illusions is the possibility of multiple partial matches. That is, representation of *no critics* in an illusion sentence like (5b) results in a match of the [+negation] retrieval cue but not the [+c-command] retrieval cue. The existence of this partial match results in a higher probability of acceptance compared to baseline sentences like (5c). This approach treats NPI licensing as analogous to other kinds of dependency resolution like agreement illusions (Wagers et al. 2009; Jäger et al. 2017). Note, however, that the profile of these illusions is not uniform (see Parker and Phillips 2016 among others).

An alternative proposal attributes illusions to the over-application of the same kinds of indirect pragmatic licensing mechanisms (discussed in §1.1) that have been proposed to account for the acceptability of NPIs in some contexts that are not explicitly negative (Xiang et al. 2009). Under this hypothesis, the use of restrictive relative clauses in NPI illusion sentences is critical, since these modifiers generate contrastive implicatures via Gricean mechanisms (e.g. Sedivy et al., 1999). Importantly, these contrastive implicatures are globally negative – that is, for a sentence like (6b), some other set of authors is inferred to have not received acknowledgement, etc. Thus, the same pragmatic mechanism that allows for NPIs in contexts like the scope of *doubt* (see (3) above) may yield an impression of acceptability for NPI illusion sentences. Critically, the claim is not that NPI illusions result from valid negative inferences which would yield full acceptability as in (3), but rather, that the possibility of generating negative inferences results in some momentary pragmatic confusion. An explanation along similar lines is proposed in Mendia et al., (2018), who suggest that some illusions arise because a covert exhaustive operator (e.g. a silent *only*) is inferred, making this contrastive implicature an entailment.

Both of these hypotheses treat NPI illusions as a highly general phenomenon – illusions should occur (with some probability) for all sentences containing a main clause NPI and structurally irrelevant licensor. However, as Orth et al. (2020a) and our Experiments 1-5 show, this generality does not exist. Rather, NPI illusions appear to be specific to contexts where the relative clause expresses a negatively quantified meaning. This surprising finding motivates a third approach, the *scalar alternatives hypothesis*, which we propose here. This hypothesis places focus on the role of scalar alternatives in NPI licensing, and predicts that illusions occur when an NPI appears shortly after a clause evoking appropriate alternatives, which we argue is more likely to occur for clauses containing negative quantifiers. This hypothesis is partly motivated by claims about the processing of negation, a literature that has proceeded largely independently from work on NPI licensing.

Researchers such as Tian & Breheny (2016) and Xiang et al. (2020) have proposed that the comprehension of a negated statement requires the activation of a Question Under Discussion (QUD), or a possible question to which the negated statement could be an answer. In most natural communicative scenarios, the common ground establishes a QUD before the negative sentence is uttered, and the negative sentence can be easily understood. Negative sentences in isolation often result in processing challenges because of the need to rapidly accommodate a QUD. For example, an out-of-the-blue utterance such as *I didn't cheat on the exam* may lead to suspicion that the speaker did in fact cheat (or has been accused of cheating). In this framework, this suspicion arises because the listener infers that the QUD is whether cheating occurred, indicating that this was not a settled part of the common ground. Thus, a negated sentence triggers consideration of the QUD to which it is an answer, as well as the other possible answers to that QUD (the alternatives). We suggest that different negative elements may evoke different QUDs – specifically, negative quantifiers may evoke scalar QUDs (and scalar alternatives) whereas simple verbal negation typically evokes polar QUDs (and binary alternatives). For example, *no one cheated on the exam* may evoke the QUD *how many people cheated on the exam?* whereas *I didn't cheat on the exam* may evoke the QUD *did the speaker cheat on the exam?*. Furthermore, scalar approaches to NPI licensing, as we discussed previously, suggest that NPIs are licensed when contained in a proposition that represents the strongest claim in a scale of pragmatic alternatives (e.g. Chierchia 2006; Fauconnier 1975a, 1975b). Combining this idea with the QUD framework just discussed, it is possible that encountering a negative quantifier such as *no* may provide an appropriate pragmatic context for NPI interpretation even before an NPI is encountered. Thus, if alternatives that have been activated within the relative clause are not rapidly inhibited, the main-clause NPI may be interpreted with respect to this representation, resulting in the fleeting perception of acceptability.

We now turn to a fourth hypothesis, initially proposed by de-Dios-Flores et al. (2017) and later developed by Orth et al. (2020a), which must be acknowledged in light of the contrast between quantificational and non-quantificational licensors that will be demonstrated in what follows. This account, which we will call the *scope miscalculation hypothesis,* treats NPI illusions as a result of the parser's failure to accurately calculate the scope of negative quantifiers. This proposal appeals to the fact that quantifiers are sometimes compatible with multiple scope interpretations, and effectively treats the acceptance of the main clause NPI as merely a side effect of a prior error in determining the interpretation of a negative quantifier. We directly address this possibility in Experiment 4.

Lastly, we note an initially appealing hypotheses that warrants brief consideration. This account proposes that an error in signal detection generates confusion of *never* for *ever*, due to the orthographic and phonological similarities of the two words. Crucially, substituting *never* in place of *ever* would provide a grammatical continuation for NPI illusion sentences. Yet, despite the appealing simplicity of such an account, findings by de-Dios-Flores (2019) show that continuations with *never* are judged unacceptable in online and offline tasks. This means that if, in NPI illusion sentences*, ever* were being mistaken for *never,* we would expect a penalty in acceptability ratings, rather than a boost, contrary to the NPI illusion pattern.

1.3 The present investigation

Here we test a few of the clearest predictions made by the hypotheses being considered as explanations for the NPI illusion. First, the memory-based hypothesis is committed to an NPI-licensing operation that retrieves licensors from prior linguistic context using some set of features that all licensors share. We can be confident that verbal negation such as *not* or *didn't* would have the relevant features, since verbal negation licenses all NPIs in English. Thus, the hypotheses clearly predicts that embedded verbal negation should cause illusions. In Experiments 1-3 we test this prediction and find that it is not borne out. Orth et al. (2020a) report similar findings. We note that the pragmatic rescuing hypothesis is also unable to straightforwardly account for this finding, whereas both the scope miscalculation hypothesis and the scalar alternatives hypothesis predict this effect.

Next, in Experiment 4, we address the predictions of the scope miscalculation hypothesis. Using a sentence comprehension task, we evaluate whether NPI illusion sentences are interpreted in a way that suggests errors in assigning quantifier scope. This hypothesis predicts high rates of negative interpretations of the main clause for sentences with embedded negative quantifiers, regardless of whether an NPI is later encountered. It additionally predicts that acceptance of illusion sentences occurs only when negative interpretations of the main clause have been established. Our findings do not align with these predictions.

Finally, in Experiments 5 and 6 we address a prediction of the scalar alternatives hypothesis. This hypothesis highlights the meanings of NPI-licensing contexts, rather than the properties of NPI licensors themselves. These last experiments address the role of meaningful content in the relative clause other than the licensor. We find that illusions can be found for verbal negation if one takes care to match the clause-level meaning to that of typical illusion sentences with negative quantifiers.

2 Contrasting licensor types

We now turn to the question of whether all non-c-commanding licensors interfere with NPI licensing. Uniformity of the illusion is a clear prediction of the memory-based hypothesis, whereas other explanations like the scalar alternatives hypothesis and the scope miscalculation hypothesis allow for variability across licensors. Indeed, the initial motivation for this manipulation was the suspicion that, contrary to verbal negation, the negative quantifier *no* implies strong or exceptionless claims, which are the kind of contexts in which NPIs are typically used. To test whether the illusion is in fact uniform across licensors, we compared standard illusion sentences like (6a) to sentences like (6b), in which the illusory licensor is non-quantificational.

(6) a. *The authors [that **no** critics recommended for the award] have **ever** received...
b. *The authors [that the critics did **not** recommend for the award] have **ever** received...
...acknowledgement for a best-selling novel.

Experiment 1 used untimed acceptability ratings in order to verify that our stimuli are appropriate. Experiments 2 and 3 used speeded acceptability judgments to determine whether sentences like (6a) and (6b) behave similarly with respect to illusions.

2.1 Experiment 1: offline acceptability

In order to establish that native speakers' perception of the experimental materials was as expected when given ample time, we conducted an offline acceptability experiment. The materials used in this and the following experiments were adapted from Parker & Phillips (2016) by adding a fourth condition with verbal negation. Additional minor edits were made to about half of stimuli in order to improve naturalness, match sentence length across stimuli, and remove modals, due to possible variability in the interaction of different forms of negation with the modal. Based on Parker & Phillips' untimed acceptability ratings, we expected to obtain a clear pattern of grammatical sensitivity: high acceptability ratings for sentences in which the NPI *ever* is licensed by a c-commanding negation and low ratings for sentences that lack a structurally-appropriate licensor. No differences were expected among the three ungrammatical conditions.

2.1.1 Participants

16 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received $2 as compensation. In this and the following experiments participants were recruited using Amazon Mechanical Turk. In order to ensure that the participants were native speakers of English they were asked to complete a native speaker qualification test and only participants that answered at least 7 out of 9 questions correctly were allowed into the task. We excluded from our analyses workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a one-sided t-test. All 16 participants met these criteria. Average ratings for grammatical fillers were 6.07 out of 7, with a standard deviation of 0.59, and average ratings for ungrammatical fillers were 3.27, with a standard deviation of 0.85.

2.1.2 Materials

The experimental materials consisted of 36 sets of items across 4 conditions that varied the presence, location, and type of licensor with respect to the NPI *ever*. This manipulation resulted in the four experimental conditions shown in Table 1. Conditions A, B, and D correspond to the standard grammatical baseline, embedded-*no*, and ungrammatical baseline conditions, respectively. Condition C (embedded-*not*) uses a similar structure to condition C in that it contains a negative word that is structurally irrelevant to the NPI, but in this condition we use verbal negation.

Table 1: Experimental materials for Experiment 1.

A. Grammatical baseline	**No** authors [that the critics recommended] have **ever** received…
B. Embedded *no*	The authors [that **no** critics recommended] have **ever** received…
C. Embedded *not*	The authors [that the critics did **not** recommend] have **ever** received...
D. Ungrammatical baseline	The authors [that the critics recommended] have **ever** received...
	...acknowledgement for a best-selling novel.

Each participant rated 108 sentences: 36 experimental items and 72 fillers of similar length and complexity. The experimental items were distributed across 4 lists using a Latin Square design and the fillers were the same in each list. Of the 72 total fillers, 42 were constructed to include a range of violations in order to encourage full use of the 1 to 7 scale. The remaining 30 filler sentences were grammatical. During the delivery of the instructions, participants were asked to complete 6 practice items to ensure that they had understood the procedure.

2.1.3 Procedure

The sentences were presented using Ibex Farm and the presentation order was randomized for each participant. The instructions asked participants to rate the sentence's acceptability using a 7-point scale in which 7 was the most acceptable value and 1 the least acceptable. Each sentence was displayed on the screen together with the scale, and participants could take as much time as needed before providing their rating. The task was completed in less than 20 minutes.

2.1.4 Analysis

The results for this and the following experiments were analyzed using a linear mixed-effects model whose maximal structure was initially built including by-subject and by-item random intercepts and slopes for the experimental conditions. When this model failed to converge, it was reduced according to the recommendations provided by Barr et al. (2013). *P*-values were derived using the lmerTest package (Kuznetsova et al. 2017). Further details are included in Supplementary Files. The relevant contrasts for this experiment used dummy coded variables as follows: to identify the effect of grammaticality, the grammatical baseline condition was used as the reference level, resulting in contrasts between this and the three ungrammatical conditions (i.e. embedded-*no*, embedded-*not*, and the ungrammatical baseline). To identify the effect of embedded negation, the ungrammatical baseline condition was used as the reference level, resulting in contrasts between the ungrammatical baseline and embedded-*no,* and the ungrammatical baseline and embedded-*not.*

2.1.5 Results

The results from this experiment are presented in Figure 1. The model results revealed a clear effect of grammaticality shown by significant differences between the grammatical baseline condition and the other three experimental conditions (grammatical baseline vs. embedded-*no*: β=-2.88, SE=0.26, t=-10.93, p<.001; grammatical baseline vs. embedded-*not*: β=-3.49, SE=0.25, t=-14.01, p<.001; grammatical baseline vs. ungrammatical baseline: β=-3.42, SE=0.28, t=-12.19, p<.001). Furthermore, ungrammatical sentences containing embedded-*no* were rated significantly higher than ungrammatical baseline sentences, though this effect was numerically small: on average, ratings of 1.76 vs 2.30 (β=0.54, SE=0.22, t=2.37, p=.03). No differences were observed between sentences containing embedded-*not* and ungrammatical baseline sentences (β=-0.07, SE=0.11, t=-0.65, p=.52).

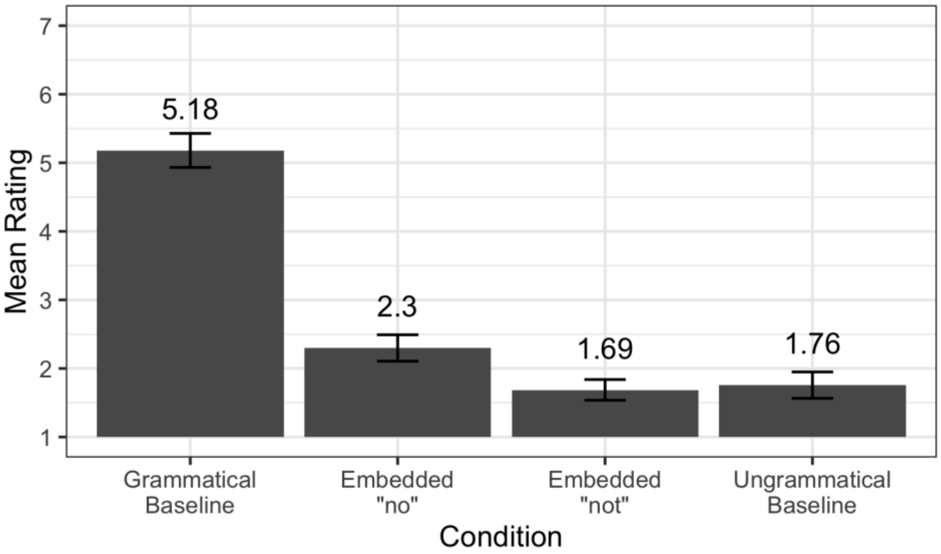

Figure 1. Mean ratings for the experimental conditions in Experiment 1. Error bars indicate standard error of the mean across subjects.

2.1.6 Discussion

The results show that participants clearly identify the grammatical baseline condition as acceptable and the ungrammatical baseline condition as unacceptable. In addition, the ratings observed for the two conditions containing embedded negative elements are highly degraded relative to the grammatical baseline. In this regard, these results confirm that, in an untimed task, speakers are sensitive to NPI licensing contrasts in our materials.

Nonetheless, we observe a small boost for the embedded-*no* condition when compared to the ungrammatical baseline condition. Note that similar patterns have been observed in other NPI illusion experiments (e.g. Xiang et al. 2006; Yanilmaz & Drury 2018). There are a few possible explanations for this boost. One possibility is that our embedded-*no* sentences are truly more acceptable than ungrammatical baseline sentences (and embedded-*not* sentences). This could be, in principle, either due to a difference in grammatical status or a difference in any other component of sentence processing that contributes to acceptability (i.e. pragmatic naturalness). Since attributing this boost to a difference in grammatical status would amount to challenging the overwhelming generalization that NPIs can only be licensed when in the scope of a licensor, we do not pursue it further. However, it is possible that our items were slightly more natural in the embedded-*no* condition compared to the other ungrammatical conditions. The other possibility is that we are observing an illusion, even in an untimed task. Note that the fact that we allow participants as much time as they need to arrive at a confident judgment does not guarantee that they will take that time. In fact, Mechanical Turk workers are strongly incentivized to complete tasks as quickly as possible. Finally, it is of course possible that we are simply observing noise and the statistically significant (*p*=.03) finding is a false positive. We are unable to determine definitively whether the cause of the boost is a difference in naturalness, a persistent illusion, or a false positive, but due to the small effect size and the fact that embedded-*no* sentences are overwhelmingly judged unacceptable, it is appropriate to proceed with these items in a speeded-acceptability task.

2.2 Experiment 2: speeded acceptability

In this and the following experiment we use a speeded acceptability judgement task to investigate the contrast between embedded quantificational negation and verbal negation in the emergence of illusion effects. Following previous speeded acceptability findings, we expect to replicate the well-documented illusion for sentences containing embedded quantificational negation *no.* The key question here is whether sentences containing embedded non-quantificational forms of negation – in this case, verbal negation *not* – in the relative clause also generate illusion effects and, particularly, if these effects are similar to those obtained with quantificational negation.

2.2.1 Participants

35 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received $3 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. Note that while this is a relatively weak threshold for inclusion, all participants also completed a native speaker questionnaire before beginning the task. Thus, our goal in excluding participants was merely to identify those who were not attending to the task and clicking randomly. 4 workers were excluded based on these criteria, resulting in 31 participants in our analysis. The mean filler-trial accuracy of the included participants was 77%.

2.2.2 Materials

The materials used in this task were the same 36 sets of experimental items and 72 filler sentences that were used in Experiment 1. Participants saw 3 practice trials before beginning the experiment.

2.2.3 Procedure

In the speeded acceptability task, each sentence was displayed word by word at a rate of 400 ms. per word, in the center of the screen, using the RSVP paradigm. At the end of each sentence participants were asked to provide a *yes*/*no* button press judgment in response to the question “Was that a good sentence?” within 2 seconds. If participants failed to provide the judgment in time, a message indicated that they were too slow. The dependent measure was the acceptance rate across trials and participants. Although sentence-final judgments are relatively late for probing incremental representations, this method has been reliably used in prior studies of NPI illusions as well as other varieties of grammatical illusions (e.g. Drenhaus et al. 2005; Parker & Phillips 2016; Parker 2019; Orth et al. 2020a). There are at least two main advantages to this method: first, the effect size for illusions is big relative to more implicit measures like reading times, and second, data can be easily collected over the internet. Note that while methods like self-paced reading and eye-tracking while reading are sometimes preferred because they allow for detailed information about the time-course of sentence processing, in this case, timing information is not necessary to answer questions like whether illusions occur. In our task, participants were instructed to read the sentences carefully and judge whether they came across as well-formed English. The task lasted for approximately 30 minutes and the order of presentation for experimental and filler sentences was randomized for each participant.

2.2.4 Analysis

Results were analyzed using the same strategy as in Experiment 1 but using logistic mixed-effects models. Experimental manipulations were coded using dummy coding, treating the ungrammatical baseline condition as the reference, resulting in three factors which compared the effects of licensing (ungrammatical baseline vs grammatical baseline), the presence of embedded-*no* (ungrammatical baseline vs embedded-*no*), and the presence of embedded-*not* (ungrammatical baseline vs embedded-*not*), respectively. An additional model was constructed to compare the embedded-*no* and embedded-*not* conditions, using the embedded-*not* condition as a reference.

2.2.5 Results

The results from this experiment are presented in Figure 2, which shows the proportion of *yes* responses given to each condition. An effect of grammaticality was observed (β=5.33, SE=0.39, z=13.70, p<.001), indicating that the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. An effect of embedded-*no* was observed (β=1.06, SE=0.28, z=3.74, p<.001), replicating the standard illusion effect for negative quantifiers. No effect of embedded-*not* was observed (β=0.26, SE=0.30, z=0.87, p=.39). Bayes factor and equivalence test analyses demonstrating that this null effect is reliable can be found in Supplementary Files. An additional model comparing embedded-*not* and embedded-*no* revealed a significant effect of the type of embedded negation (β=0.85, SE=0.28, z=3.01, p=.003), indicating that the embedded-*no* condition was significantly more likely to be judged acceptable than the embedded-*not* condition.

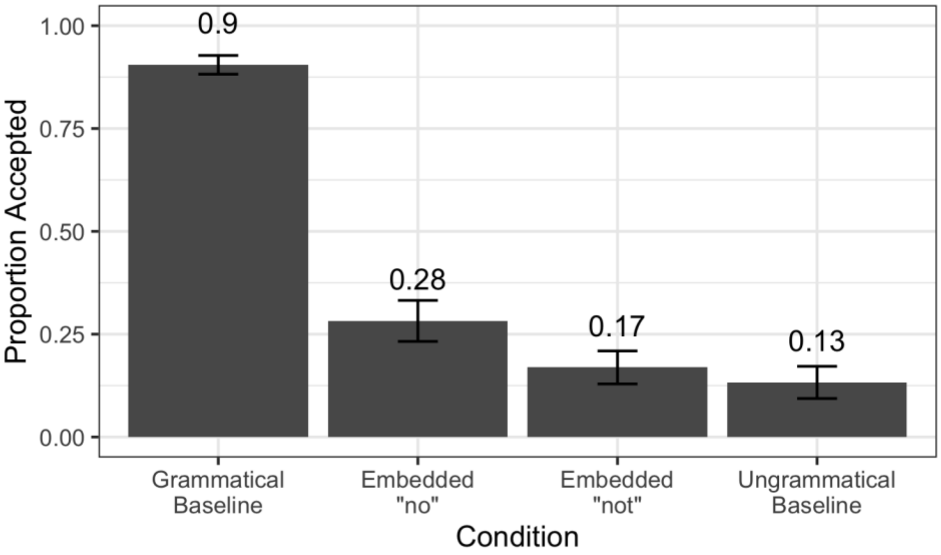

Figure 2. Mean percentage of 'YES' responses for the experimental conditions in Experiment 2. Error bars indicate standard error of the mean across subjects.

2.2.6 Discussion

The results from Experiment 2 replicate the widely attested illusion effect for sentences containing *no* as an embedded licensor. This pattern of illusory licensing is not found when the embedded licensor is *not*. Before discussing the implications of this finding in detail, it is necessary to consider two potential confounds that could be behind the observed results. First, it is possible that participants' judgements across the task are influenced by comparisons made across trials. Since our experiment included both the quantificational negation and verbal negation conditions, it is in principle possible that illusions do exist for verbal negation, but these illusions are suppressed in our experiment due to cross-trial comparisons. A parallel issue has to do with the fact that sentences containing the non-contracted form of verbal negation (*did not* as opposed to *didn't*) might be perceived as less natural. The lower acceptability ratings we observed for this condition could be due to that unnaturalness, instead of a lack of illusions. We address both of these issues in the next experiment.

2.3 Experiment 3: speeded acceptability

In order to address the concerns stated above, some small modifications were introduced to the materials and the design. First, the critical illusion conditions (embedded-*no* and embedded-*not*) were tested as a between-subjects factor to avoid possible complications due to cross-trial comparisons. This results in two sub-experiments, each of which included the same grammatical and ungrammatical baseline conditions and only one of the embedded-licensor conditions. Second, in the verbal negation condition we used a contracted form to increase naturalness.

2.3.1 Participants

49 participants were recruited for this task. The first group (24 participants) performed the sub-experiment with embedded-*no*. The second group (25 participants) performed the sub-experiment with embedded-*haven't*. All participants provided informed consent and they

received $3 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more and workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test. 1 worker was excluded from our analyses for the *not* sub-experiment resulting in 24 participants per sub-experiment (48 participants total). The mean filler-trial accuracy of the included participants was 81%.

2.3.2 Materials

The experimental materials used in this task consisted of 36 sets containing the same four experimental conditions as Experiments 1 and 2. The experimental materials used here introduced some minor changes with respect to the previous experiments (cf. Table 2). As already noted above, we used a contracted form of verbal negation to increase naturalness. We also used the present perfect in the relative clause (where previously we had used the simple past), so that all four conditions had the same structure and the same number of words.

Table 2: Experimental materials for Experiment 3.

A. Grammatical baseline	**No** authors [that the critics have recommended in their reviews] have **ever**…
B. Embedded *no*	The authors [that **no** critics have recommended in their reviews] have **ever**…
C. Embedded *haven't*	The authors [that the critics **haven't** recommended in their reviews] have **ever**...
D. Ungrammatical baseline	The authors [that the critics have recommended in their reviews] have **ever**...
… received acknowledgement for a best-selling novel.	

For the *haven't* sub-experiment, we selected the grammatical baseline, ungrammatical baseline, and embedded-*haven't* versions of our 36 items, and created three lists from these, together with 90 filler sentences of similar internal structure, length and complexity. Similarly, we constructed three lists for the *no* sub-experiment from the same set of items. The number of filler items was slightly increased with respect to the previous experiments in order to have more variability within the task. Grammaticality was balanced so that approximately half of the sentences were ungrammatical across the task. Each list had a total of 126 items and participants were randomly assigned to one of the six lists. During the delivery of the instructions, participants were asked to complete 6 practice items.

2.3.3 Procedure

The speeded acceptability procedure followed the same steps as in Experiment 2. The task lasted for approximately 35 minutes.

2.3.4 Analysis

Results were analyzed in the same way as Experiment 2. Experimental manipulations were coded using dummy coding, treating the ungrammatical baseline condition as the reference, resulting in two factors which compared the effects of licensing (ungrammatical baseline vs grammatical baseline) and the presence of an embedded licensor (ungrammatical baseline vs embedded-negation) respectively. Models comparing the two sub-experiments used sum coding for the factor comparing them.

2.3.5 Results

The results are shown in Figure 3. A main effect of grammaticality was observed (β=4.05, SE=0.20, z=19.76, p<.001), indicating that, averaging across sub-experiments, the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. A main effect of embedded-negation was observed (β=0.58, SE=0.17, z=3.35, p<.001), indicating that, averaging across sub-experiments, the embedded-negation conditions were significantly more likely to be judged acceptable than the ungrammatical baseline condition. We additionally observed a significant negation-type by sub-experiment interaction (β=-0.67, SE=0.17, z=-3.84, p<.001). Planned comparisons indicated that the effect of embedded-negation was present only for the *no* sub-experiment (β=1.28, SE=0.24, z=5.44, p<.001). The *haven't* sub-experiment revealed no significant effect of embedded-negation (β=-0.08, SE=0.26, z=-0.33, p=.75). Once again, because the claims rely on a null finding, we present analyses using bayes factors and equivalence tests in Supplementary Files.

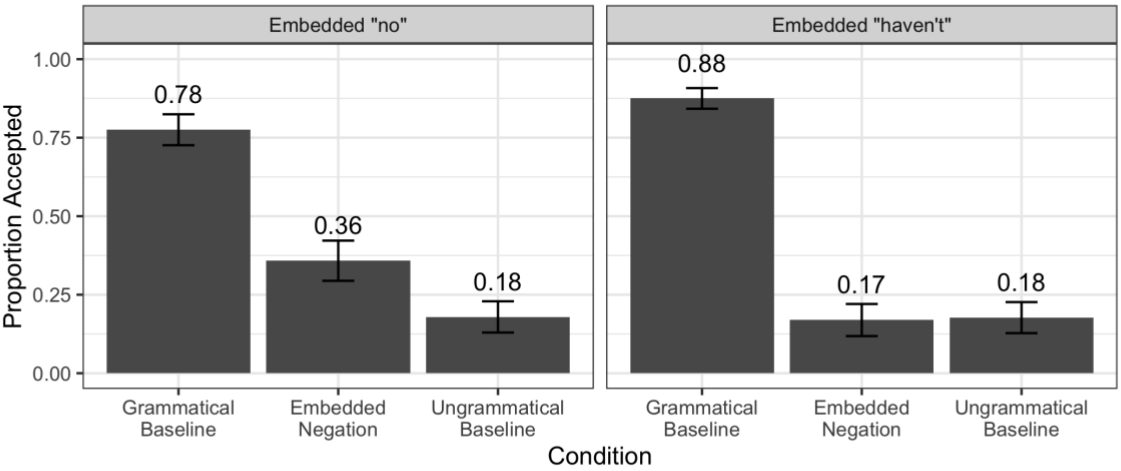

Figure 3. Mean percentage of 'YES' responses for the experimental conditions in Experiment 3. Error bars indicate standard error of the mean across subjects.

2.3.6 Discussion

The results from Experiment 3 replicate those from Experiment 2 in all relevant aspects. First, there is a clear illusion effect for sentences containing embedded-*no*. Second, this task replicated the lack of statistically significant illusion effects for sentences containing embedded non-quantificational negation. Since the results from Experiments 2 and 3 align, we conclude that the use of uncontracted negation *not* or the presentation of the two embedded conditions together did not have a critical influence in the results of Experiment 2. What is certain so far is that NPI illusions are not general across all forms of embedded negation. We now evaluate the four hypotheses presented in section 1.2 in light of these findings.

First, the memory-based hypothesis (Vasishth et al. 2008), straightforwardly predicts the erroneous retrieval of embedded licensors to be uniform across *no, not,* and *haven't,* because both all of these should share the relevant features in their encodings (i.e. [+negation]). Any attempt to adapt the feature set to capture the observed difference (e.g. adding [+quantificational]) will lead to inappropriate predictions in cases of true licensing, since both *no* and *not/haven't* are perfectly capable of licensing an NPI within their scope. Our findings thus suggest that NPI illusions are not a consequence of erroneous retrievals of a partially feature-matching licensor in memory. Secondly, recall that the pragmatic rescuing hypothesis (Xiang et al. 2009; Xiang et al. 2013) proposes that negative contrastive inferences are responsible for illusions. For example, sentences like (6a), a standard NPI illusion configuration, license the inference in (6b), which, under this hypothesis leads to increased acceptance of NPIs within *P*. However, note that the exact same inference is supported by sentences like (7a), which includes non-quantificational negation *haven't* in the relative clause. Thus, the hypothesis predicts uniformity in the illusion across at least these two licensors, contrary to our findings.

(6) a. The authors [that no critics recommended] have P
b. The authors [that the critics HAVE recommended] have NOT P

(7) a. The authors [that the critics haven't recommended] have P
b. The authors [that the critics HAVE recommended] have NOT P

We now turn to the predictions of the scalar alternatives hypothesis. Recall that this hypothesis places blame for illusions on the brief persistence of scalar alternatives following the conclusion of the relative clause, and the NPI's use of these alternatives as a licensing context. Thus, the hypothesis predicts a contrast between quantificational and non-quantificational forms of negation to the extent that these forms trigger different types of alternatives to the relative clause. Recall that independent research on the processing of negation highlights the importance of a negated sentence's relation to a QUD, and we suspect that not all forms of negation will be appropriate for the same type of QUD. In particular, sentences containing a negative quantifier like *no* may be well suited to scalar QUDs, with alternative answers corresponding to other quantifiers that could be substituted in place of *no*. While other forms of negation like *not* or *haven't* are also perfectly compatible with scalar alternatives (as this is precisely the type of alternatives they contrast with when used with a truly licensed NPI), we suggest that this is not the default interpretation assigned to verbal negation in the absence of an NPI or a supporting context. If these assumptions are correct, then it is the discrepancy between the inferred alternatives for quantificational and non-quantificational negation that drives the contrast in their impact as illusory licensors observed in Experiments 1-3. These assumptions also predict that there will be overlap in the types of contexts that make negative quantifiers appropriate (in virtue of the QUDs they answer) and the types of contexts that make NPIs appropriate (in virtue of their licensing conditions). Because non-quantificational negation can be used in response to a wider variety of QUDs, we expect to see less overlap between these contexts and NPI-containing contexts.

We conducted a corpus search in order to determine whether there is in fact more overlap between uses of quantificational negation and NPIs than uses of non-quantificational negation and NPIs. We drew 5000 random instances of sentences containing quantificational negation (*no*) and another 5000 random instances of sentences containing verbal negation (*-n't* or *not*) from the Corpus of Contemporary American English (Davies 2008). They were manually analyzed for the presence of the NPI *ever* at any sentence position in the scope of the negative element. The results showed that environments with *no* were five times more likely to also contain *ever* (59, 1.18%) than environments with *not*/*n't* (12, 0.24%). The difference was statistically significant ($X^2(1)= 31.34$, $p<0.001$). Note that this is not a claim that comprehenders track the surface statistics of the co-occurance of quantificational negation and *ever* versus non-quantificational negation and *ever*, but rather both speaker and comprehender are sensitive to an underlying correspondence between certain types of communicative goals (i.e. expressing a meaning at the strong endpoint of a pragmatic scale) and the surface forms that achieve those goals. Returning to the question of why NPI illusions arise, the scalar alternatives hypothesis predicts that it is specifically the interpretive work of inferring appropriately-ordered scalar alternatives prior to the NPI that creates vulnerability when the NPI is encountered. Because verbal negation may not trigger such alternatives on its own (or may do so to a lesser extent) due to its much broader range of possible uses, verbal negation is not expected to give rise to illusions.

Finally, the contrast observed in Experiments 1-3 is also predicted by the scope miscalculation hypothesis. Under this hypothesis, a comprehension error occurs at the quantifier, in the computation of its scope, and the acceptance of the NPI is a downstream consequence of this prior error. Errors in NPI licensing are not predicted for non-quantificational licensors because these licensors are not vulnerable to errors in scope computation, under this hypothesis. In Experiments 4-6 we evaluate further predictions of this hypothesis and do not find support for it.

3 The interpretation of NPI illusions

While there is substantial prior work (including our Experiments 1-3) manipulating the factors that cause illusions, very little is known about the interpretation a reader comes away with after experiencing an illusion. This is an understandable gap, since probing interpretations is methodologically difficult and it is not always clear that the hypotheses being considered make clear predictions about the interpretation, so interpretation measures may not be scientifically useful. However, the scope miscalculation hypothesis does in fact make predictions about comprehenders' sentence-final interpretations, and so we attempted to correct this data gap in the present experiment.

3.1 Experiment 4: interpretation and acceptability

Recall that the key claim of the scope miscalculation hypothesis is that NPI illusions arise not due to any problems in the processing of NPIs per se, but due to problems in the representation of the scope of the licensor, as if it takes scope over the whole main clause, not just the relative clause in which it occurs. Once this prior error has been made, the NPI can be easily licensed by the negative element, which, in the comprehender's internal representation, does take scope over it. Under this hypothesis, verbal negation is invulnerable to this scope error because only quantifiers have such a wide set of possible scopal interpretations. It is worth noting that although some quantifiers, in some contexts, do have the possibility to take scope from positions other than their surface form, negative quantifiers do not routinely exhibit this variability (Liu 1990), and quantifiers do not in general take scope out of a relative clause. Thus, the high-scope representation that is entertained by the comprehender on some trials under this account is truly an error – a "scope illusion" – as it is not a representation that is licensed by the grammar. This of course begs the question of why parsing is not faithful to the grammar in the case of quantifier scope processing (whereas parsing is faithful to the grammar in the case of NPI processing, under this hypothesis), but this is beyond the scope of the current investigation.

Here we address a straightforward prediction of the scope miscalculation hypothesis: if the error in processing is in fact an error that occurs at the quantifier, and if acceptance of the main clause NPI is merely a side effect of this prior problem, we should see interpretive consequences of this error (i.e., globally negative interpretations) regardless of whether an NPI occurs later in the sentence. That is, after hearing sentences like (8a) or (8b), comprehenders should, under this hypothesis, come away believing that the authors in question have not received acknowledgment for their novels, because the main clause is under the scope of negation in their internal representation.

(8) a. *The authors that **no** critics have recommended have **ever** received acknowledgment
for a best-selling novel.
b. The authors that **no** critics have recommended have received acknowledgment
for a best-selling novel.

The rate of negative interpretations of these sentences should surely not be 100%. We know this because NPI illusions do not occur on every trial. But critically they should be similar regardless of the inclusion of the NPI. This is the key prediction addressed by Experiment 4.

We note that an alternative scope miscalculation hypothesis is possible, in which rather than stochastically landing on either a high-scope or low-scope representation of the quantifier prior to the NPI, the comprehender maintains, in parallel, both scope representations until at least the NPI position. The NPI then is taken as evidence that the high-scope representation was the intended one, because only this representation allows the NPI to be licensed. We will refer to these as the early-scope-assignment and late-scope-assignment versions of the scope miscalculation hypothesis, respectively. This version of the scope miscalculation hypothesis does not predict equal rates of globally negative interpretations for (8a) and (8b). Both scope miscalculation hypotheses, however, predict that negative interpretations and acceptance of the NPI should go hand in hand, since they are both consequences of the same representational error, they differ only in the timecourse of when that representation is selected. That is, both predict that the NPI illusion trials that are interpreted globally negatively should be the trials that are accepted, and the NPI illusion trials that are interpreted globally positively should be the trials that are rejected. We therefore asked participants for both a comprehension question judgment and an acceptability judgment on every trial, in order to determine whether this correspondence exists.

3.1.1 Participants

We recruited 33 participants using Amazon Mechanical Turk. Workers received $10 for completion of the 45-minute task. We excluded workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test, as well as those who failed either the “instructions quiz” (see “Procedure” section below) or the “attention checks” (see “Materials” section below). 4 workers were excluded from our analyses resulting in 29 participants. The mean filler-trial accuracy of the included participants was 79% for acceptance judgments and 92% for interpretation judgments.

3.1.2 Materials

This experiment probed interpretations by asking participants directly about the asserted content of the main clause. The comprehension questions for experimental trials were always written such that a *no* answer indicated that the participant believed that the main clause assertion was under the scope of negation and a *yes* answer indicates that the participant believes that the main clause assertion is not under the scope of negation. The 36 item sets from Experiment 3 were used to generate the eight conditions shown in Table 3. This included the four conditions with *ever* with various licensor positions, as well as four conditions which are identical except that the NPI *ever* has been omitted. These four NPI-free conditions allow us to determine whether interpretations corresponding to erroneous quantifier scope assignment arise independent of the NPI. We also included 72 fillers, which were balanced both for *yes* and *no* interpretation responses and for grammaticality. Before beginning the experiment, participants judged six practice trials. We also included eight attention check trials, which were intended to identify participants who were clicking answers randomly without reading the sentences. For example, the sentence for an attention check trial might have read “After this sentence, please choose 'no' as the answer.” It was then followed by only one question, “What is the answer?”. Participants who responded incorrectly to two or more attention check trials were excluded from our analyses (see “Participants” section above).

Table 3: Experimental materials for Experiment 4.

With *ever*	**A. Main clause *no***	**No** authors that the critics recommended have **ever** received...
	B. Embedded *no*	The authors that **no** critics recommended have **ever** received...
	C. Embedded *didn’t*	The authors that the critics **didn't** recommend have **ever** received...
	D. Without negation	The authors that the critics recommended have **ever** received...
Without *ever*	**E. Main clause *no***	**No** authors that the critics recommended have received...
	F. Embedded *no*	The authors that **no** critics recommended have received...
	G. Embedded *didn’t*	The authors that the critics **didn't** recommend have received...
	H. Without negation	The authors that the critics recommended have received...
		...acknowledgment for a best-selling novel.

3.1.3 Procedure

Participants viewed sentences presented one word at a time with rapid serial visual presentation, just as in Experiments 2 and 3. At the conclusion of each sentence, participants first gave an untimed binary judgment of acceptability, and then were asked a comprehension question about the main clause. For example, the sentences in Table 2 were all followed by the question “Have the authors received acknowledgement for a novel?”. The instructions were careful to instruct participants to interpret definite descriptions as referring to the same set of individuals as was mentioned in the sentence (i.e. *the authors* in the question refers to the same authors that were discussed in the sentence), so that the questions could be stated without the inclusion of relative clauses in the questions themselves. The possible responses were *yes*, *no*, and *I can’t answer*. The *I can’t answer* option was to be used in cases where the participant found the sentence not only ungrammatical but uninterpretable. There was no time limit for responses. Because this task was somewhat more complex than previous experiments, we also included a brief quiz following the instructions to ensure that participants had read and understood the instructions correctly. For example, participants were asked “When should you choose 'I can't answer'?” and had to choose from the options (A) “Whenever I don't have an opinion on the question”, (B) “Whenever the sentence was ungrammatical”, or (C) “When I can't determine what the sentence meant”. The correct answer was displayed after participants made their selection. Participants who responded incorrectly to two or more task quiz questions were excluded from our analyses (see “Participants” section above).

3.1.4 Analysis

Trials receiving an *I can’t answer* comprehension question response were simply removed from our analyses (both acceptability data and comprehension data). This amounted to 82 trials across our eight experimental conditions (7.8% of the total number of experimental trials). Results were analyzed using logistic mixed effects models, as in Experiments 2 and 3. For analyses of comprehension data, we are interested in the probability of a negative interpretation, so responses were coded as 1 in the case that comprehenders gave a *no* response and 0 in the case that comprehenders gave a *yes* response. Experimental manipulations were coded using dummy coding, treating the ungrammatical baseline condition as the reference, unless otherwise stated.

3.1.5 Results

Overall results are shown in Figure 4. We focus first on the NPI illusion data – that is, acceptance rates for the four conditions that included *ever*. An effect of grammaticality was observed (β=8.05, SE=3.00, z=2.68, p=.007), indicating that the main-clause-*no* condition was significantly more likely to be judged acceptable than the without-negation condition. An effect of embedded-*no* was observed (β=1.73, SE=0.58, z=2.98, p=.003), replicating the standard illusion effect for negative quantifiers. No effect of embedded-*didn't* was observed (β=0.14, SE=0.42, z=0.34, p=.74).

Having established that the NPI illusion (and lack of illusions for embedded-*didn't*) replicates in this task, we turn to the key predictions of the scope miscalculation hypothesis. To determine whether there are “scope illusions”, we analyzed negative interpretation rates for the four conditions without *ever*. An effect of the presence of main-clause-*no* was observed (β=10.00, SE=4.05, z=2.47, p=.01), indicating that the main-clause-*no* condition was significantly more likely to be interpreted negatively than the without-negation condition. An effect of embedded-*no* was observed (β=1.91, SE=0.87, z=2.19, p=.03), indicating that embedded quantifier sentences are more likely to be interpreted negatively than the without-negation condition. No effect of embedded-*didn't* was observed (β=0.02, SE=1.01, z=0.01, p=.98). Thus, we do observe statistically reliable "scope illusions" in approximately 15% of trials.

To evaluate the key prediction of the early-scope-assignment version of the hypothesis, we ask whether there is an impact of the presence of *ever* on negative interpretation rates for embedded-*no* sentences. A model comparing negative interpretation rates for only these two conditions (embedded-*no* with *ever* vs. embedded-*no* without *ever*) revealed an effect of the presence of *ever* (β=4.36, SE=0.83, z=5.24, p<.001). This finding indicates that embedded-quantifier sentences containing *ever* were in fact substantially more likely to be interpreted negatively than embedded-quantifier sentences not containing *ever*.

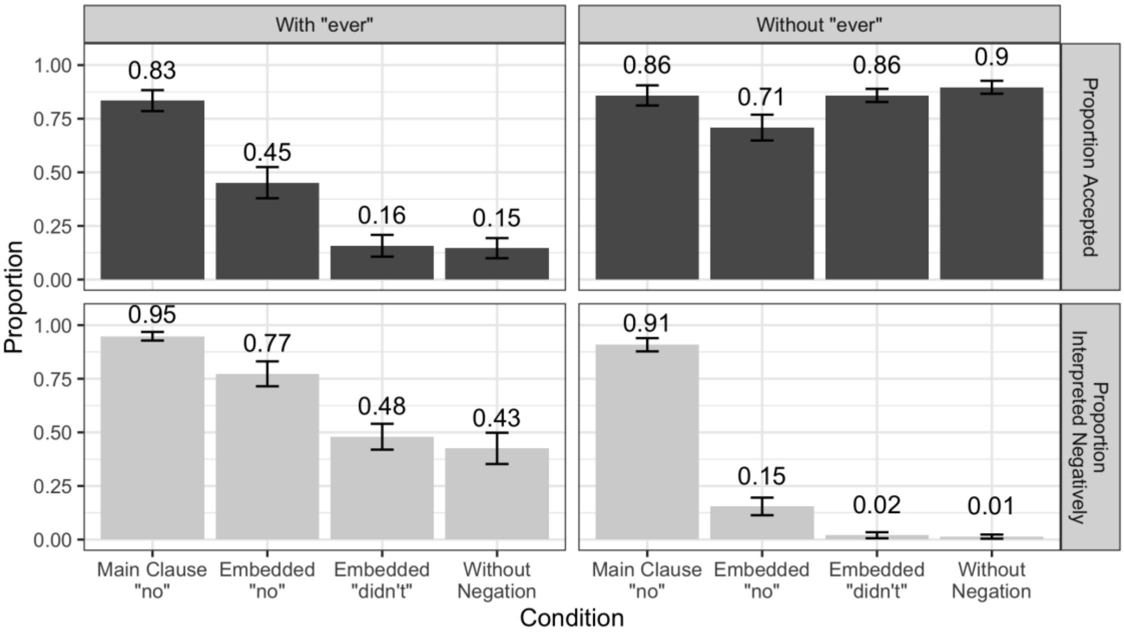

Figure 4: Mean percentage of 'YES' responses for the experimental conditions in Experiment 4 in the acceptability judgement (top) and mean percentage of 'NO' responses for the experimental conditions in Experiment 4 in the interpretation question. Error bars indicate standard error of the mean across subjects.

Lastly, we evaluate whether globally negative interpretations and NPI acceptance go hand-in-hand as both versions of the scope hypothesis predict. We evaluate this prediction by focusing just on the three standard illusion conditions (that is, conditions with *ever*, looking at just the grammatical baseline, ungrammatical baseline, and embedded-*no* conditions), and splitting the data by interpretation (*yes* vs *no* answers to comprehension questions). The relevant conditions are displayed in Figure 5. We constructed a model using the ungrammatical baseline condition as a reference, and using sum coding for trial-level interpretations (*yes* or *no* answers to comprehension questions). A main effect of grammaticality was observed (β=4.84, SE=1.81, z=2.68, p=.007), indicating that, averaging across positively- and negatively-interpreted trials, grammatical baseline sentences were more likely to be accepted than ungrammatical baseline sentences. A main effect of embedded-*no* was observed (β=1.32, SE=0.59, z=2.22, p=.026), indicating that, averaging across positively- and negatively-interpreted trials, embedded-*no* sentences were more likely to be accepted than ungrammatical baseline sentences (i.e., the NPI illusion was observed). However, no effect of interpretation was observed (β=0.13, SE=0.69, z=0.19, p=.85), and interpretation did not interact with either the grammaticality effect (β=1.70, SE=1.81, z=0.94, p=.35) or, critically, the illusion effect (β=0.67, SE=0.98, z=0.69, p=.49). Thus, we see no evidence that illusions are specific to negatively-interpreted trials.

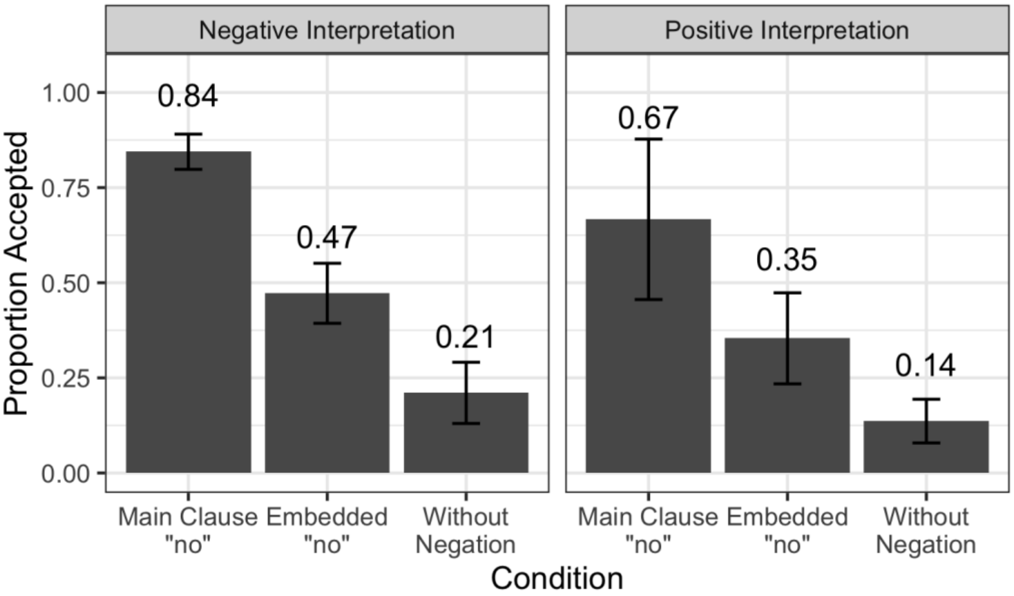

Figure 5: Mean percentage of 'YES' responses for the experimental conditions in Experiment 4 in the acceptability judgement as a function of interpretation: negative (left) and positive (right). Error bars indicate standard error of the mean across subjects.

3.1.6 Discussion

Experiment 4 was designed to evaluate whether the interpretation of NPI illusion sentences is consistent with the predictions of the scope miscalculation hypothesis. In short, our findings demonstrate the following: (a) a replication of previous findings of an illusion for embedded-*no* and lack of illusions for embedded-*didn't*, (b) while there are some negative interpretations for sentences with embedded quantifiers, these are much less frequent for sentences without NPIs than those with NPIs, and (c) the NPI illusion does not appear to be specific to negatively-interpreted trials.

Findings (b) and (c) present a challenge for the scope miscalculation hypothesis. Recall that the early-scope-assignment version of the hypothesis in essence claims that NPI illusions are merely a side effect of a previous error – a "scope illusion", in which the negative quantifier is represented as taking scope over the main clause. However, finding (b) demonstrates that globally negative interpretations seem to be a consequence of NPI illusions, not a cause of them. Additionally, both the early- and late-scope-assignment versions of the hypothesis predict that, among NPI illusion sentences, the trials that are accepted should be the ones that are interpreted negatively and the trials that are rejected should be the ones that are interpreted positively, contrary to finding (c). It is important to acknowledge that finding (c) is a null effect, since we fail to observe a statistically significant interaction. It is perfectly possible that we simply have not achieved adequate statistical power to detect an effect that is truly there. However, both the hypothesis predicts not only that illusion rates should be higher for negatively-interpreted trials, but that illusion trials arise exclusively in negatively interpreted trials, and are nonexistent in positively interpreted trials. Thus, a statistical interaction is a low bar, relative to the predictions of the hypothesis, but the data do not support even this. Additionally, we believe there are independent reasons to disfavor the late-scope-assignment version of the hypothesis. Critically, because this model allows for a delay in scope assignment until at least the NPI position, it predicts that a PPI in the same main clause position would never be subject to interference from embedded negative quantifiers, because the presence of a PPI would indicate that low scope is the best interpretation of the quantifier. However, reported findings from Orth et al. (2020c) contradict this prediction, showing illusions of ungrammaticality for main-clause PPIs preceded by embedded quantifiers.

In addition to the critical findings from Experiment 4 with respect to the scope miscalculation hypothesis, the fact that we replicate the illusion even when there are comprehension questions should put to rest concerns that illusions in acceptability tasks reflect a kind of "shallow processing" in which comprehenders fail to represent sentences at all levels of representation.

4. Sentential negation and scalar meanings

While the findings from Experiment 4 present evidence against the scope miscalculation hypothesis, the scalar alternatives hypothesis does not make obvious predictions for the kinds of interpretations tested here. In Experiments 5 and 6 we directly compare these two hypotheses. Specifically, we investigate whether illusions are categorically absent for sentences with embedded-*haven't* or if instead illusions are possible, but substantially reduced in magnitude. Because the scope miscalculation hypothesis attributes errors to quantifier processing per se, it should be impossible to observe an illusion with any non-quantificational licensor. In contrast, the scalar alternatives hypothesis attributes the illusion to the clause-level scalar alternatives, and therefore predicts illusions for any configuration that evokes the same alternatives. Crucially, verbal negation is not incompatible with these alternatives, it just is not associated with them by default in isolation. It must of course be possible to interpret a negative sentence with respect to scalar alternatives, as this is what negated sentences with NPIs mean. Under this hypothesis, then, we should be able to bring out a scalar interpretation of verbal negation by pairing it with an NPI inside the relative clause as in (9a).

(9) a.*The critics [that **haven't** recommended **any** authors of alternative genres] have **ever**...
b. The critics [that **haven't** recommended the authors of alternative genres] have **ever**...
...objected to mainstream literary trends.

Thus, these two hypotheses make clearly contrasting predictions. If NPI illusions are a side effect of scope illusions (as the scope miscalculation hypothesis claims), NPI illusions should be equally impossible for (9a) as for (9b). If instead, illusion magnitude varies as a function of the alternatives evoked in the RC, (9a) may yield illusions despite its use of non-quantificational negation *haven't* as an embedded licensor.

4.1 Experiment 5: offline acceptability

We first conducted an offline rating study to establish that the materials are appropriate. The materials used in this experiment were adapted from our stimuli from experiments 1-4. We expected to obtain a clear pattern of grammatical sensitivity and no differences are thus expected among the four ungrammatical conditions regardless of the presence and type of a structurally irrelevant licensor.

4.1.1 Participants

15 US-based native speakers of English participated in this experiment. All participants provided informed consent and they received $9 as compensation. We excluded workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a one-sided t-test, as well as workers who failed the “attention checks” (see “Materials” below). 1 worker was excluded from our analyses, resulting in 14. participants. Among the included participants, average ratings for grammatical fillers were 5.81 out of 7, with a standard deviation of 0.53, and average ratings for ungrammatical fillers were 3.66, with a standard deviation of 0.82.

4.1.2 Materials

The experimental materials for this and the following tasks consisted of 40 sets of 5 items. The inclusion of a condition like example (13) which includes *not* followed by *any* in order to match the clause-level meaning of the condition with *no*, required that we convert our stimuli from object relative clauses to subject relative clauses. Note that merely changing the clause type would result in strange meanings for many stimuli (for example, our standard example, *the authors that no critics recommended* would become *the authors that recommended no critics*, which is inconsistent with world knowledge). This required further modification to improve naturalness and plausibility. A sample set of the five experimental conditions is shown in Table 4.

Table 4: Sample set of experimental materials for Experiment 5.

A. Gram. baseline	**No** critics [that have recommended any authors of alternative genres] have **ever**...
B. Embedded *no*	The critics [that have recommended **no** authors of alternative genres] have **ever**...
C. Embedded *not any*	The critics [that **haven’t** recommended **any** authors of alternative genres] have **ever**...
D. Embedded *not*	The critics [that **haven’t** recommended the authors of alternative genres] have **ever**...
E. Ungram. baseline	The critics [that have recommended the authors of alternative genres] have **ever**...
...objected to mainstream literary trends.	

Each participant was asked to rate 130 sentences: 40 experimental items and 90 fillers of similar length and complexity. The experimental items were distributed across 5 lists using a Latin Square design and the fillers were the same in each list. Participants completed two practice items before beginning the task. We again included eight attention check trials, randomly interspersed through the experiment. As in Experiment 4, the purpose was to identify and remove participants who chose ratings randomly without having read the sentence. For example an attention check trial read "For this sentence, please choose six as the answer." Participants who answered two or more attention check trials incorrectly were excluded from our analyses (see "Participants" section above).

4.1.3 Procedure

The offline acceptability procedure followed the same steps as in Experiment 1.

4.1.4 Analysis

The results for this experiment were analyzed using a linear mixed-effects model as in Experiment 1. Experimental manipulations were coded using dummy coding, treating the grammatical baseline condition as the reference, resulting in four factors which reflected effects of grammaticality. To determine the effect of embedded licensors, an additional model used the ungrammatical baseline as the reference level, providing the relevant contrasts between the ungrammatical baseline and embedded-*no*, embedded-*haven't-any,* and embedded-*haven't* conditions respectively.

4.1.5 Results

The results from this experiment are presented in Figure 1. Linear mixed effects models revealed a clear effect of grammaticality shown by significant differences between the grammatical baseline condition and the other four experimental conditions (grammatical vs. embedded-*no*: β=-2.42, SE=0.18, t=-13.18, p<.001; grammatical vs. embedded-*haven't-any*: β=-2.69, SE=0.20, t=-13.16, p<.001; grammatical vs. embedded-*haven't*: β=-2.86, SE=0.18, t=-15.90, p<.001; grammatical vs. ungrammatical baseline: β=-2.79, SE=0.19, t=-14.99, p<.001). We again compared the embedded-negation conditions to the ungrammatical baseline and again found a small (average ratings of 2.60 vs 2.96) but statistically significant boost for embedded-*no* (β=0.36, SE=0.17, t=2.09, p=.04) but no such difference for embedded-*haven't-any* (β=0.09, SE=0.15, t=0.60, p=.55) or embedded-*haven't* (β=-0.07, SE=0.14, t=-0.50, p=.62).

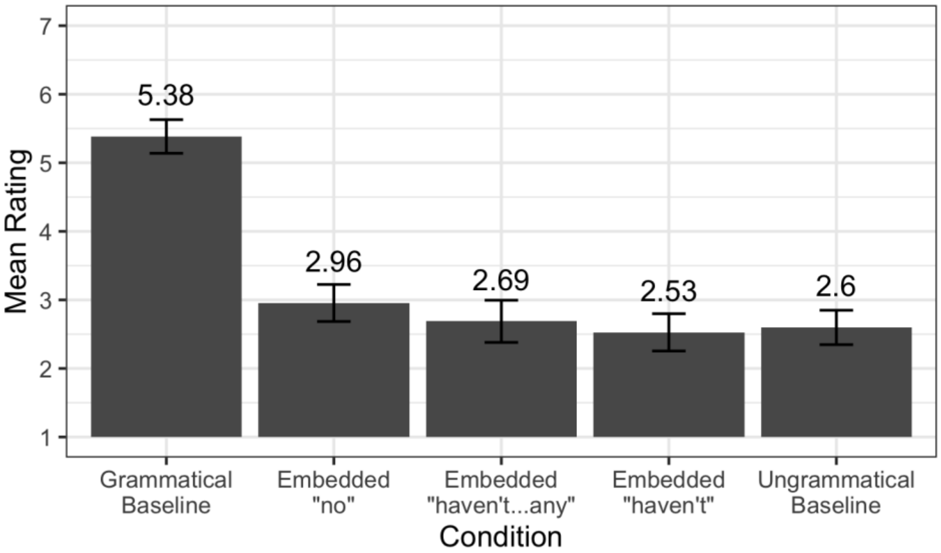

Figure 6. Mean ratings for the experimental conditions in Experiment 5. Error bars indicate standard error of the mean across subjects.

4.1.6 Discussion

The main objective of this experiment was to confirm the grammatical status of the experimental materials. The results show that participants clearly identify the grammatical baseline condition as acceptable and the ungrammatical baseline condition as unacceptable. In addition, the ratings observed for the three conditions containing non-c-commanding negative elements were highly degraded relative to the grammatical baseline. In this regard, these results provide a confirmation that speakers are sensitive to NPI licensing contrasts in our materials. Note that we again observe a numerically small but statistically significant boost in ratings for embedded-*no*, even in an offline task (see discussion of Experiment 1).

4.2 Experiment 6: speeded acceptability

We used speeded acceptability measures with the goal of determining whether sentences containing embedded-*haven't* followed by *any* inside the relative clause yield illusions. The scope miscalculation hypothesis predicts that the *haven't-any* condition should pattern with the *not* condition, because both conditions are missing the critical ingredient for illusions: a negative quantifier. In contrast, the scalar alternatives hypothesis predicts that the *haven't-any* condition could pattern with the *no* condition because the critical ingredient is the clause-level scalar meaning, which in this case has been matched.

4.2.1 Participants

195 US-based native speakers of English participated in this experiment. Note that we increased the sample size for this experiment. This was done for two reasons. First, the offline data in Experiment 5 suggest that the switch to subject relative clause may result in less clear judgments of acceptability – that is, even the grammatical and ungrammatical baseline conditions were judged closer to the middle of the scale. Thus, a larger sample was necessary to maintain adequate statistical power for a measurement with increased noise. Second, because the critical finding is now not only whether illusions arise for each of three embedded-negation conditions, but how those conditions compare to each other, the effect size of interest may be smaller (as

corroborated by pilot data). All participants provided informed consent and they received $6 as compensation. We excluded workers who failed to provide a response within 2 seconds for 25% of fillers or more, workers whose judgments of filler trials did not reliably distinguish between grammatical and ungrammatical fillers, based on a chi-squared test, and workers who failed the “attention checks”. 40 workers were excluded based on these criteria, resulting in 155 participants. The mean filler-trial accuracy of the included participants was 80%.

4.2.2 Materials

The materials used in this task were the same 40 sets of experimental items and 90 filler sentences that were used in Experiment 5. Participants completed two practice trials before beginning the experiment. We additionally included eight attention check trials, as in Experiments 4 and 5. Participants who answered two or more attention check trials incorrectly were excluded from our analyses (see “Participants” section).

4.2.3 Procedure

The speeded acceptability procedure was identical to Experiments 2 and 3.

4.2.4 Analysis

Results were analyzed using logistic mixed effects models. Experimental manipulations were coded using dummy coding, treating the ungrammatical baseline condition as the reference, resulting in four factors which compared, respectively, the effects of licensing (ungrammatical baseline vs grammatical baseline), the presence of embedded-*no* (ungrammatical baseline vs embedded-*no*), the presence of embedded-*haven't-any* (ungrammatical baseline vs embedded-*haven't-any*), and the presence of embedded-*haven't* (ungrammatical baseline vs embedded-*haven't*). An additional model was constructed using the embedded-*haven't* condition as a baseline in order to compare embedded-*haven't* to embedded-*no*. Finally, a third model compared the embedded-*haven't-any* condition to the two other embedded-negation conditions, using the *haven't-any* condition as the reference.

4.2.5 Results

The results from this experiment are presented in Figure 2, which shows the percentage of *yes* responses given to each condition. An effect of grammaticality was observed (β=4.58, SE=0.25, z=18.19, p<.001), indicating that the grammatical baseline condition was significantly more likely to be judged acceptable than the ungrammatical baseline condition. An effect of embedded-*no* was observed (β=0.60, SE=0.12, z=4.74, p<.001), replicating the standard illusion effect for negative quantifiers, though the effect size is numerically somewhat smaller than we observed in Experiments 2, 3, and 4 (β values of 1.06, 1.28, and 1.73, respectively, compared to 0.60 here). A statistically significant but again small effect of embedded-*haven't* was also observed (β=0.31, SE=0.11, z=2.80, p=.001). Note that this is inconsistent with our findings in Experiments 2, 3, and 4 which did not reveal statistically reliable illusions for sentences with embedded-*not/-n’t*. Lastly, a statistically significant and again small effect of embedded-*haven't-any* was also observed (β=0.43, SE=0.11, z=3.90, p<.001).

An additional model comparing embedded-*haven't* and embedded-*no* revealed a significant effect of the type of embedded negation (β=0.29, SE=0.12, z=2.34, p=.02), indicating that the embedded-*no* condition was significantly more likely to be judged acceptable than the embedded-*haven't* condition, replicating our findings of a contrast between *no* and *not* from Experiments 2, 3, and 4. Finally, a third model comparing the embedded-*haven't-any* condition to the two other embedded negation conditions did not identify statistically significant differences between either embedded-*haven't-any* and embedded-*haven't* (β=-0.12, SE=0.11, z=-1.12, p=.26) or embedded-*haven't-any* and embedded-*no* (β=0.17, SE=0.12, z=1.39, p=.17). That is to say, although embedded-*no* and embedded-*haven't* differed from one another, embedded-*haven't-any* was numerically intermediate and statistically not distinguishable from either one.

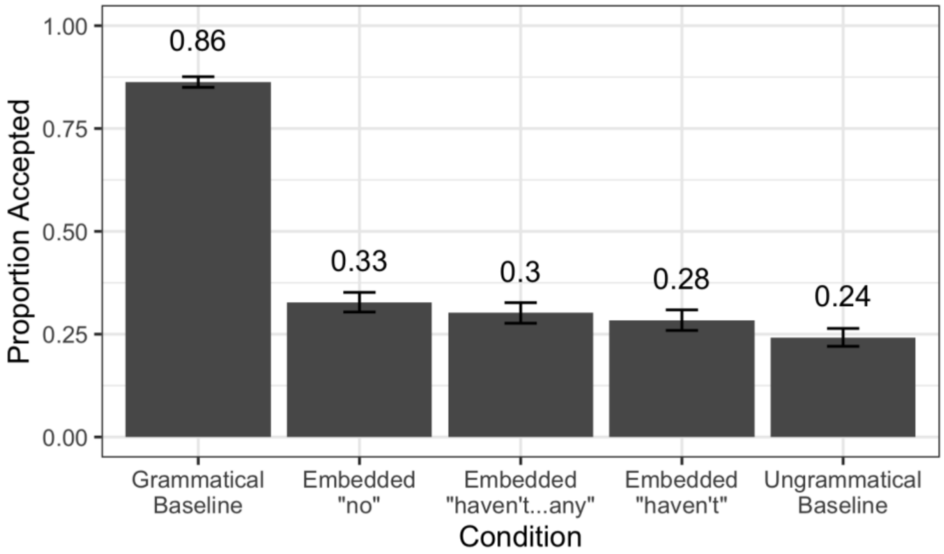

Figure 7. Mean percentage of 'YES' responses for the experimental conditions in Experiment 6. Error bars indicate standard error of the mean across subjects.

4.2.6 Discussion

In order to determine whether the lack of illusions for embedded-*not/-n't* in prior experiments was due to the categorical impossibility of illusions with non-quantificational embedded licensors or instead merely a (substantial) reduction in the probability of an illusion due to the wide range of possible non-scalar uses of verbal negation, we measured illusion rates for both the previously-tested embedded-*no* and embedded-*haven't* as well as a novel condition containing embedded-*haven't* paired with NPI *any* inside the relative clause. While we successfully replicated three important contrasts – those between grammatical vs ungrammatical baseline, embedded-*no* vs ungrammatical baseline, and embedded-*no* vs embedded-*haven't*– we additionally found, to our surprise, statistically reliable illusions for embedded-*haven't.* Recall that our question was whether embedded-*haven't-any* would pattern with embedded-*no* and yield illusions or pattern with embedded-*haven't* and fail to yield illusions. However, we found that all three embedded negation conditions yielded illusions, and embedded-*haven't-any* could not be statistically distinguished from either of the other two conditions.

There are essentially three potential reasons we could have observed an illusion for embedded-*haven't* in this experiment despite having observed a lack of illusions for embedded-

not/-n't in Experiments 2, 3, and 4: (a) the true state of the world is that there are no illusions for embedded-*not/-n't* and the observed effect in Experiment 6 was a false positive, (b) the true state of the world is that there are illusions for embedded-*not/-n't* and the observed null effects in Experiments 2, 3, and 4 were false negatives (potentially due to insufficient statistical power), or (c) illusions sometimes arise for embedded-*not/-n't,* and they are sensitive to some factor that we inadvertently manipulated between Experiments 2-4 and Experiment 6. We address each of these possibilities in turn.

If the observed illusion for embedded-*not/-n't* in Experiment 6 is a false positive, the take-aways are clear: we observed an illusion for embedded-*haven't-any*, contrary to the predictions of the scope miscalculation hypothesis and consistent with the predictions of the scalar alternatives hypothesis.

In order to address the possibility that the lack of illusions for embedded-*not/-n't* in Experiments 2, 3, and 4 were false negatives, we conducted power analyses. We estimate the effect size of the illusion for embedded-*not/-n't* at β=0.3, as this was the observed effect size in Experiment 6. Power analysis conducted using simr (Green & MacLeod 2016) revealed that Experiments 2, 3, and 4 had only between 15% and 20% power to detect an effect size of β=0.3. Note that this does not mean that Experiments 2-4 were simply "under-powered experiments". They were designed to detect illusion effects of the size typically observed for embedded-*no* (for which the observed β is routinely above 1.0), and for this they were adequately powered. Thus, we cannot rule out the possibility that there are illusions for embedded-*not/-n't* with a small (β=0.3) effect size. Of course, this does not impact the claim that there is a contrast between embedded-*no* and embedded-*not/-n't.* As discussed above, a key difference between the scope miscalculation hypothesis and the scalar alternatives hypothesis is that the first of these treats the possibility of illusions as categorical: scope illusions (of which NPI illusions are a side effect) under this hypothesis happen for quantificational licensors and they do not happen for non-quantificational licensors. In contrast, under the scalar alternatives hypothesis, the illusion rate can be turned up or down as a function of the likelihood of scalar inferences in the relative clause. Thus, if there are small (β=0.3) illusions for embedded-*not/-n't* and larger (β=0.6 to β=1.0+) illusions for embedded-*no,* this would again be consistent with the scalar alternatives hypothesis.

Finally, we address the possibility that illusions for embedded-*not/-n't* arose in Experiment 6 due to some inadvertently manipulated factor. One notable change was the shift to subject relative clauses instead of object relative clauses in Experiment 6. However, both Orth et al. (2020c) and AUTHORS (to be submitted) used subject relative clauses and observed a lack of illusions for embedded-*not/-n't* and so we do not think it likely that this is the key difference driving the observed illusions for embedded-*not/-n't* in Experiment 6. A more promising candidate for the critical change is the inclusion of the *haven't-any* condition. It is possible that over the course of the experiment, participants adjust their interpretations based on other stimuli they have seen. Under the scalar alternatives hypothesis one such adjustment might be to interpret clauses containing *not* with respect to scalar alternatives. If this were the case, we would expect the earliest trials to show a cleaner pattern of results - illusions for embedded-*no* and embedded-*haven't-any* but not for embedded-*haven't.* The results from Experiment 6 for only the first fifth of the experiment (i.e., the first 28 trials per participant) are given in Figure 8. While this is of course a post-hoc exploratory analysis, it does appear to be the case that in early trials, illusions were observed for embedded-*no* and embedded-*haven't-any* but not for embedded-*haven't*. This pattern is consistent with the predictions of the scalar alternatives hypothesis.

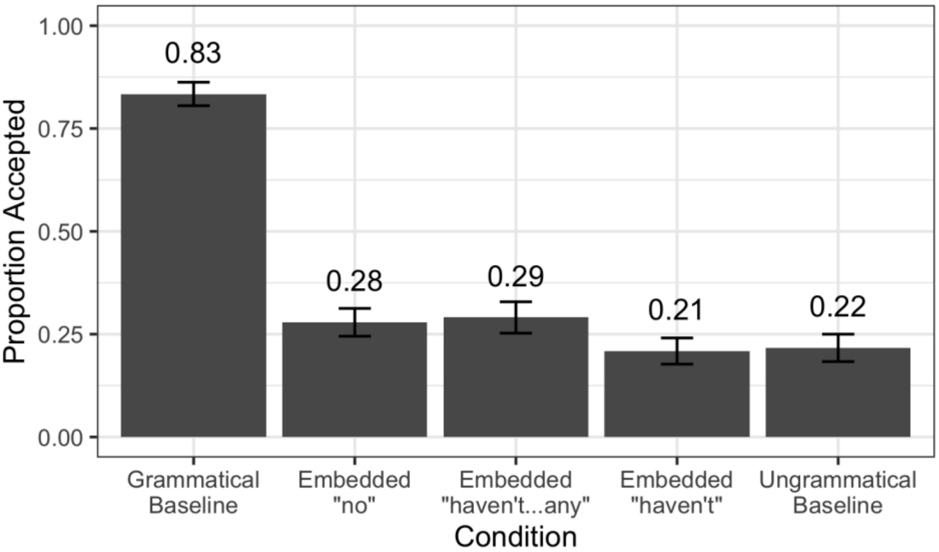

Figure 8: Mean percentage of 'YES' responses for the experimental conditions the first fifth of Experiment 6. Error bars indicate standard error of the mean across subjects.

In sum, while we cannot definitively determine the cause of the sudden appearance of an illusion for embedded-*haven't* in Experiment 6, we find that a number of possible explanations for this finding are all more consistent with the claims of the scalar alternatives hypothesis than with those of the scope miscalculation hypothesis.

5 General discussion

5.1 Key findings

The evidence presented here provides a strong case for narrowing down the existing range of hypotheses on NPI illusions. In Experiments 2, 3, and 4 we established a clear contrast in the illusion profile for sentences containing embedded-*no* such as (15) compared to sentences containing embedded-*not/-n't* such as (16), as has also been shown in Orth (2020a). Based on these three Experiments, it appears that illusions arise only for embedded negative quantifiers. However, in light of our Experiment 6 findings, we might more accurately say that illusions are substantially reduced (though not necessarily absent) for sentences like (10b).

(10) a. *The authors [that **no** critics recommended for the award] have **ever** received acknowledgement for a best-selling novel.
b. *The authors [that the critics did **not** recommend for the award] have **ever** received acknowledgement for a best-selling novel.

Experiment 4 additionally investigated the sentence-final interpretation of illusion sentences by asking participants for both an acceptability judgment and a response to a question like "Did the authors receive acknowledgements for their novels?" following sentences like (10a) and (10b). These findings reveal that although NPI-illusion sentences like (10a) are often understood as expressing a globally negative meaning, this interpretive error only arises once the NPI is encountered. The interpretive error is therefore not a possible cause of problems in processing the NPI, but rather a consequence.

Experiment 6 measured illusion rates for sentences like (11a-c) in order to determine whether illusions are categorically impossible for non-quantificational embedded licensors (*haven't* in b and c) or if the illusion rate can be turned up or down as a function of the clause's broader meaning. The results tentatively suggest that clause-level meanings play an important role, though we acknowledge that the surprising observation of an illusion for embedded-*haven't* in this experiment complicates the interpretation of the results.

(11) a. *The critics [that have recommended **no** authors of alternative genres] have **ever**...
b. *The critics [that **haven't** recommended **any** authors of alternative genres] have **ever**...
c. *The critics [that **haven't** recommended the authors of alternative genres] have **ever**...
...objected to mainstream literary trends.

Thus, the key critical empirical contributions are as follows: (a) verbal negation in the form of *not* or *–n't* yields few or no illusions, (b) erroneous globally-negative interpretations are not independently established prior to the NPI, and (c) aspects of the relative clause meaning beyond just the presence of a negative word, appear to play a role in illusions.

We additionally propose a novel explanation for the NPI illusion, in the form of the scalar alternatives hypothesis. This hypothesis assumes that the online computation of NPI licensing is not the retrieval of a negative word from memory, but rather the integration of the NPI meaning into a clause-level meaning, which is only successful when the appropriate scalar alternatives are available. The key error, then, is merely the failure to rapidly inhibit scalar alternatives to the relative clause prior to the NPI (see Muller et al. submitted for further exploration of the narrow window of vulnerability). This hypothesis attributes the contrast between *no* and *not* as embedded licensors to differences in the alternatives they evoke. It additionally predicts that when embedded-*not/-n't* is contained in a scalar-alternatives-evoking clause it should behave similarly to *no* with respect to illusions, which we find tentative support for in Experiment 6. We note, however, that while the hypothesis does not make especially strong predictions with respect to sentence-final interpretation, our Experiment 4 finding that majority of illusion trials are interpreted as if the main clause were negative is somewhat surprising under this account. If the problem resulting in the erroneous acceptance of the NPI is that the relative clause and its alternatives are still being considered as a local context, it is then not clear why the main clause's meaning is so drastically altered. We suspect that this sentence-final globally-negative interpretation is a result of reanalysis processes that attempt to reconcile an acceptable NPI with a syntactic parse that clearly represents the NPI in the main clause. Very little is known about the processes that take place from the point the NPI is encountered until a decision is made, and more work is clearly needed to determine whether this is a plausible explanation for the interpretation findings we obtain.

5.2 Alternative explanations for NPI illusions

We additionally considered three other explanations for the NPI illusion. The memory-based hypothesis (Vasishth et al. 2008) attributes illusions to partial matches in memory between a set of search cues ([+negative] and [+c-command], though other feature sets are possible) and the features encoded on chunks in memory representing prior words and phrases. While this hypothesis is appealing in its generality, capturing NPI illusions, agreement attraction, and other illusion phenomena under the same mechanism, it fails to account for the specificity of the NPI illusion. That is, this hypothesis cannot explain the observed contrast between *no* and *not,* since these licensors would be encoded with the same [+negation] feature which should allow them to interfere with NPI processing.

Similarly, the pragmatic rescuing hypothesis (Xiang et al. 2013) seems to predict uniformity across embedded licensors, contrary to our findings. Recall that this hypothesis attributes illusions to erroneous inferences from (12a) to (12c), but these inferences seem equally available for relative clauses containing non-quantificational negation as in (12b).

(12) a. The authors [that no critics recommended] have P
b. The authors [that the critics haven't recommended] have P
c. The authors [that the critics HAVE recommended] have NOT P

In fact, even the ungrammatical baseline sentences used in NPI illusion experiments (21a) could license negative contrastive inferences that would license an NPI in P (13b), since the inference arises in virtue of the restrictive relative clause, not the negative form inside the relative clause.

(13) a. The authors [that the critics recommended] have P
b. The authors [that the critics have NOT recommended] have NOT P

To address this, Xiang et al. (2009) suggest that "speakers may be more likely to generate such inferences if the contrasting referents are made very salient in the discourse. Negative quantifiers can do exactly this" (Xiang et al. 2009: 53). If they are correct in asserting that negative quantifiers are critical, this could in principle explain the contrast between embedded-*no* and embedded-*not/-n't*. To support the idea that negative quantifiers make contrasting referents salient, Xiang and colleagues point to work showing that following a downward entailing quantifier, comprehenders readily accept reference with a pronoun to the complement set of the quantifier (Sanford et al. 1996). For example, *they* in (14) (examples from Sanford et al. 1996) can be understood as referring to the many football fans who did not go to the game, rather than the few who did attend (the referent of *few of the football fans* in the first sentence).

(14) Few of the football fans went to the game. They watched it on TV instead.

However, it is not clear how the possibility of reference to the complement set of a quantifier is related to the inference from (12a) to (12c), which is the critical inference for NPI illusions, under Xiang et al.'s (2013) hypothesis. The mechanism explored by Sanford and colleagues predicts that following a sentence like (15), we may expect reference to the (many) critics who declined to recommend the authors (i.e., the complement set to the downward-entailing quantifier-phrase *few critics*). This mechanism does not generate inferences to any other set of authors, only to other sets of critics. Thus, we do not find a clear prediction of a contrast between quantificational and non-quantificational forms of negation under the pragmatic rescuing hypothesis.

(15) The authors [that few critics recommended] have P

Finally, we consider two variants of a hypothesis that attributes NPI illusions to problems in correctly assigning quantifier scope, which was suggested by both de-Dios-Flores et al. (2017) and Orth et al. (2020a). The core claim of the quantifier scope hypothesis is that the negative quantifier is represented as if it takes scope over the entire main clause; given this scope assignment the NPI can be licensed. Non-quantificational forms of negation would not give rise to illusions due to their more limited scope possibilities, thus capturing the observed contrast between embedded-*no* and embedded-*not/-n't.* Note, however, that negative quantifiers are also quite limited in their scope possibilities (Liu 1990), and so the wide-scope representation that gives rise to the illusion requires a parsing procedure that is willing to temporarily ignore some scope-related constraints of the grammar. Why the scope constraints of negative quantifiers are violable but the scope constraints of other forms of negation are not violable is not obvious.

The two relevant variants of this hypothesis are one in which scope assignment is early, such that on some trials a wide-scope representation has already been established prior to the NPI, and a version in which multiple scope configurations are considered in parallel until at least the NPI position, at which point a choice is made, taking the NPI as evidence for a wide-scope representation. If scope assignment occurs prior to the NPI, we would expect to see interpretive consequences for the main clause regardless of whether an NPI is eventually encountered, contrary to our findings in Experiment 4. Under both versions of the hypothesis, we would expect the NPI illusion trials that are interpreted negatively to be the ones that are accepted and the trials that are interpreted positively to be the ones that are rejected. This is not what we observe in Experiment 4, but note that the finding relies on the lack of statistical significance of a two-way interaction. Both versions of the hypothesis also predict that illusions are impossible for non-quantificational forms of embedded negation, contrary to our findings in Experiment 6.

5.3 Aligning parsing and grammar

We have addressed a number of competing grammatical hypotheses that aim to account for the distribution of NPIs in natural language, as well as a number of competing processing hypotheses that aim to account for the error profile of the comprehension of NPI-containing sentences. While we do not aim to choose between grammatical accounts on the basis of processing data, some brief discussion of the possible alignment between grammar and parsing is warranted. It is of course not logically necessary that the representations that guide incremental interpretation are identical to those that are licensed by the mental grammar. Notable two-system views include Bever's "quick-and-dirty" parsing strategies (1970) and Ferreira and colleagues' "good enough hypothesis" (2002). While it may be tempting to treat grammatical illusions as an obvious case of two-system processing (since the defining characteristic of illusions is the divergence between initial representations and grammatically-licensed representations), Parker (2019) provides modeling evidence from agreement attraction showing that changes over time in the perceived grammatical status of an incoming sentence are expected under both two-system and one-system accounts. In the case of NPIs we have argued for an account of illusions which relies on the present sentence's relation to scalar alternatives, invoking a particular style of grammatical explanation – namely, those in the spirit of Fauconnier (1975a, 1975b), which emphasize a critical role between the NPI and the context in which it appears. This does not, of course, rule out the possibility that an alternative grammatical proposal is preferable and the computation of scalar alternatives is a "shortcut" that allows for rapid online processing, but it is not obvious how this mechanism is more rapid or more efficient than a direct implementation of the grammar.

6 Conclusion

This work proposes an alternative explanation that emphasizes the relevance of real-time semantic interpretation – specifically, the availability of scalar inferences – for the emergence of NPI illusions. This claim is largely motivated by the finding that illusions are robust for negative quantifiers but reduced or absent for non-quantificational forms of embedded negation. Importantly, this finding calls into question the previously promising hypothesis that a wide range of linguistic illusions can be explained by the properties of the memory architecture. While it is clear that memory systems are critical to language comprehension, it appears that the re-framing of NPI licensing as merely the retrieval of a prior lexical item in memory is both unfaithful to hypothesized grammars of NPIs and inconsistent with our results. We additionally review some competing hypotheses including the erroneous pragmatic licensing hypothesis and scope miscalculation hypothesis and find that they do not predict the error profile nor the interpretation patterns that we observe. The scalar alternative hypothesis proposed here allows for NPI illusions to result from the same detailed operations that are deployed during routine NPI dependency resolution. We think that this is a useful shift in our thinking about how grammatical knowledge is deployed by the parser.

Data Availability/Supplementary Files

The experimental materials, results and analysis scripts (including the supplementary analyses) for all the experiments and the corpus data can be found in the following repository: https://osf.io/nt79x/?view_only=9211706032064b1fa5614a65c5b71e3a

Competing interests

The authors have no competing interests to declare.

References

AUTHORS (to be submitted). The time-course of NPI illusions.

Baker, C. L. 1970. Double Negatives. *Linguistic Inquiry*. The MIT Press 1(2). 169–186.

Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the development of language*. New York: Wiley. 279–362.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the ‘logicality’ of language. *Linguistic Inquiry* 37. 535–590. https://doi.org/10.1162/ling.2006.37.4.535

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): One billion words, 1990-2019. www.english-corpora.org/coca.

de-Dios-Flores, Iria. 2019. Processing Sentences With Multiple Negations: Grammatical Structures That Are Perceived as Unacceptable. *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.02346

de Dios-Flores, Iria, Hanna Muller & Colin Phillips. 2017. Negative polarity illusions: licensors that don't cause illusions, and blockers that do. Poster presented at the CUNY Conference on Human Sentence Processing, Boston.

Drenhaus, Heiner, Douglas Saddy & Stefan Frisch. 2005. Processing negative polarity items. When negation comes through the backdoor. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*. Berlin: Mouton de Gruyter. 145–165.

Fauconnier, Gilles. 1975a. Pragmatic Scales and Logical Structure. *Linguistic Inquiry* 6(3). 353–375.

Fauconnier, Gilles, 1975b. Polarity and the scale principle. *Proceedings of the Chicago Linguistics Society* 11. 188-199.

Ferreira, Fernanda, Karl G. D. Bailey & Vittoria Ferraro. 2002. Good-enough representa- tions in language comprehension. *Current Directions in Psychological Science* 11. 11–15. https://doi.org/10.1111/1467-8721.00158

Giannakidou, Anastasia. 2006. Only, Emotive Factive Verbs, and the Dual Nature of Polarity Dependency. *Language* 82(3). 575–603. https://doi.org/10.1353/lan.2006.0136

Green, Peter & Catriona MacLeod. 2016. simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4). 493–498. https://doi.org/10.1111/2041-210X.12504

Hildebrandt, Luis & Matthew Husband. 2019. Quantifiers, restrictors, and illusory NPI licensing. Poster presented at the CUNY Conference on Human Sentence Processing, Boulder.

Israel, Michael. 1997. The scalar model of polarity sensitivity. In Danielle Forget, Paul Hirschbühler, France Martineau & María Luisa Riveiro (eds.), *Negation and polarity: Syntax and Semantics*. Amsterdam/Philadelphia: John Benjamins. 209–230. https://doi.org/10.1075/cilt.155

Israel, Michael. 2011. Grammar of polarity: Pragmatics, sensitivity and the logic of scales. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511975288

Jäger, Lena, Felix Engelmann & Shravan Vasishth. 2017. Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. Journal of Memory and Language, 94. 316--339. https://doi.org/10.1016/j.jml.2017.01.004

Kadmon, Nirit & Fred Landman. 1993. Any. *Linguistics and Philosophy* 16(4). 353–422. https://doi.org/10.1007/BF00985272

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

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(1). 1–26. https://doi.org/10.18637/jss.v082.i13

Ladusaw, William A. 1980. Polarity Sensitivity as Inherent Scope Relations. New York: Garland Publishing.

Ladusaw, William A. 1996. Negation and Polarity Items. In Shalom Lappin (ed.), *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell Reference. 321–341.

Laka, Itziar. 1994. On the syntax of negation. New York, London: Garland Publishing. https://doi.org/10.4324/9781315049465

Lee, So Young, Aydogan Yanilmaz, Jiwon Yun & John E. Drury. 2018. The processing of Turkish and Korean NPI licensing and intrusion: ERP evidence. Poster presented at the CUNY Conference on Human Sentence Processing, Davis.

Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3). 375–419. https://doi.org/10.1207/s15516709cog0000_25

Linebarger, Marcia C. 1987. Negative polarity and grammatical representation. *Linguistics and Philosophy* 10(3). 325–387. https://doi.org/10.1007/BF00584131

Liu, Feng-hsi. 1990. Scope dependency in English and Chinese. PhD dissertation, University of California. Los Angeles: University of California.

Mendia, Jon Ander, Ethan Poole & Brian Dillon. 2018. Spurious NPI licensing and exhaustification. *Proceedings of the 28th Semantics and Linguistic Theory Conference*. 233–250.

Muller, Hanna, Iria de-Dios-Flores & Colin Phillips. 2019. Not (just) any licensors cause negative polarity illusions. Poster presented at the CUNY Conference on Human Sentence Processing, Boulder.

Ng, Anne & Matthew Husband. 2017. Interference effects across the at-issue/not-at-issue divide: Agreement and NPI licensing. Poster presented at the CUNY Conference on Human Sentence Processing, Cambridge.

Orth, Wesley, Masaya Yoshida & Shayne Sloggett. 2020a. Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47(6), 906–947. https://doi.org/10.1037/xlm0000957

Orth, Wesley, Masaya Yoshida & Shayne Sloggett. 2020b. Tracking the time course of NPI illusion: Why do only some appear online? Poster presented at the 33rd Annual CUNY Conference on Human Sentence Processing. Amherst.

Orth, Wesley, Masaya Yoshida & Shayne Sloggett. 2020c. Illusions of ungrammaticality: Evidence from PPI. Poster presented at the 33rd Annual CUNY Conference on Human Sentence Processing. Amherst MA.

Parker, Dan. 2019. Two minds are not always better than one: Modeling evidence for a single sentence analyzer. *Glossa: a journal of general linguistics* 4(1). 64. https://doi.org/10.5334/gjgl.766

Parker, Dan & Colin Phillips. 2011. Illusory negative polarity item licensing is selective. Poster presented at the CUNY Conference on Human Sentence Processing, Stanford.

Parker, Dan & Colin Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition* 157. 321–339. https://doi.org/10.1016/j.cognition.2016.08.016

Sanford, Anthony J., Linda M. Moxey & Kevin B. Paterson. 1996. Attentional focusing with quantifiers in production and comprehension. *Memory & Cognition,* 24(2). 144-155. https://doi.org/10.3758/bf03200877

Sedivy, Julie C., Michael K. Tanenhaus, Craig G. Chambers & Gregory N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71(2). 109–147. http://dx.doi.org/10.1016/S0010-0277(99)00025-6

Tian, Ye & Richard Breheny. 2016. Dynamic pragmatic view of negation processing. In *Negation and polarity: Experimental perspectives*. Cham: Springer. 21–43.

Vasishth, Shravan, Sven Brüssow, Richard L. Lewis & Heiner Drenhaus. 2008. Processing polarity: how the ungrammatical intrudes on the grammatical. *Cognitive Science* 32(4). 685–712. https://doi.org/10.1080/03640210802066865

Wagers, Matthew, Ellen Lau & Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. Journal of Memory and Language 61(2). 206-237. https://doi.org/10.1016/j.jml.2009.04.002.

Xiang, Ming, Brian Dillon & Colin Phillips. 2006. Testing the strength of the spurious licensing effect for negative polarity items. Talk presented at the CUNY conference on Human Sentence Processing, New York.

Xiang, Ming, Brian Dillon & Colin Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* 108(1). 40–55. https://doi.org/10.1016/j.bandl.2008.10.002

Xiang, Ming, Julian Grove & Anastasia Giannakidou. 2013. Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology* 4. 708:1–19. https://doi.org/10.3389/fpsyg.2013.00708

Xiang, Ming, Alex Kramer & Ann E. Nordmeyer. 2020. An informativity-based account of negation complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46(10). 1857–1867. https://doi.org/10.1037/xlm0000851

Yanilmaz, Aydogan & John E. Drury. 2018a. Intervening and non-intervening Interference. Poster presented at the CUNY Conference on Human Sentence Processing, Davis.

Yanilmaz, Aydogan & John E. Drury. 2018b. Prospective NPI licensing and intrusion in *Turkish. Language, Cognition and Neuroscience* 33(1). 111–138. https://doi.org/10.1080/23273798.2017.1371779

Yun, Jiwon, So Young Lee & John E. Drury. 2018. Negative polarity illusion in Korean. In Celeste Guillemot, Tomoyuki Yoshida & Seunghun J. Lee (eds.), *Proceedings of the 13th Workshop on Altaic Formal Linguistics*. MIT Working Papers in Linguistics 88. Cambridge: MIT Press.