Introduction. Fallibility in the online implementation of negative polarity item (NPI) licensing has been taken as evidence for a key role for memory retrieval operations in sentence processing [7]. However, this approach faces two challenges: it predicts uniformity in the profile of the illusion, contrary to existing findings [6, 8], and it assumes that NPI licensing relies on a search for a c-commanding negative word, whereas most theoretical proposals for NPI licensing focus on semantic properties of contexts [3]. In experiment 1 we show that online licensing also makes reference to contexts, and further reveal that the time required to suppress irrelevant contexts is key to illusions. In experiment 2 we show that illusions only arise when the nearby context makes scalar alternatives available. Finally, experiment 3 rules out an alternative explanation for the contrast observed in experiment 2 and begins to illuminate the interpretive consequences of NPI illusions. Thus we show that some of the same key concepts that explain the semantic and pragmatic constraints on NPI licensing are implicated in illusions. We further suggest that hypotheses about the processing of negation have consequences for NPI processing, bringing together two sub-literatures that have been surprisingly disconnected.

Background. Negative polarity illusions arise in (some) sentences where an NPI co-occurs with, but is not c-commanded by a negative element (1). Such sentences are unacceptable upon careful reflection (and ungrammatical on most theories of NPI licensing) but rapidly and incrementally generated representations do not consistently make this unacceptability apparent. In tasks that tap into early representations (speeded acceptability, reading times, ERPs) these sentences are accepted more frequently than baseline sentences with ungrammatical NPIs (2), though not as frequently as truly grammatical sentences with NPIs (3) [2, 6, 8]. Here we focus on speeded acceptability judgments, as they provide a robust and straightforward measure of illusion rates.

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

(2) * The authors [that the critics recommended] have **ever** received ...

(3) **No** authors [that the critics recommended] have **ever** received ...

Experiment 1. Added material preceding the NPI has been shown to "turn off" the illusion [6], though prior work does not identify *where* material needs to be added to have this effect. Here we demonstrate that illusions are sensitive to distance from the licensing environment to the NPI; distance to the negative word is irrelevant. In experiment 1 (speeded acceptability; 44 subjects; 36 items; data collected using Amazon Mechanical Turk) we introduce identical added material inside the relative clause and outside the relative clause, and find that only material added outside the relative clause has the effect of "turning off" the illusion. These findings demonstrate that the online computation of NPI licensing makes reference to clause-level properties but not individual lexical licensors (and thus closely parallels the NPI licensing constraints proposed in the theoretical NPI literature), since it is distance from the clause not distance from the licensor that matters. We additionally take these findings to suggest that it is difficulty in rapidly inhibiting representations of the nearby relative clause that causes illusions.

Experiment 2. Prior investigations of illusions used only intrusive negative quantifiers (e.g., *no*) in their stimuli, but the memory-based hypothesis predicts that all NPI licensors should yield illusions. We find in experiment 2 (speeded acceptability; 48 subjects; 36 items; data collected using Amazon Mechanical Turk) that while intrusive negative quantifiers consistently give rise to illusions, intrusive sentential negation (e.g., 4) does not. This pattern is robust across forms of sentential negation (as in *haven't*, *did not*) and clause types (subject relative clauses, object relative clauses).

(4) * The authors [that the critics **haven't** recommended] have **ever** received ...

This finding is puzzling if all that matters for NPI licensing is a clause with the appropriate entailments (i.e., downward entailing, anti-veridical, etc.), since clauses with *no* and clauses with *not* share these properties. We can make sense of these facts by considering the incremental generation of alternatives that negative sentences support. Some of the difficulty of negation processing has been explained as a consequence of the need to rapidly infer contextual alternatives to the negated sentence (usually framed in terms of a question under discussion). Much work on NPI licensing suggests that the semantic strength of the NPI-containing sentence relative to alternatives is crucial [1, 4, 5]. We propose that clauses containing negative quantifiers make scalar alternatives available even before the NPI is encountered, thus allowing licensing to proceed more quickly, such that illusions arise. For example, *the authors that no critics recommended* evokes alternatives like *the authors that some critics recommended*, suggesting that the speaker is identifying authors as the low endpoint of some scale of quality or recommendability. Contrast this with *the authors that the critics didn't recomment*, which evokes only a binary split in authors, not necessarily a full scale. The contrast in licensors can be explained as a consequence of this contrast in the availability of scalar alternatives.

Experiment 3. An alternative explanation for the contrast we observe in experiment 2 places blame on the computation of quantifier scope. This hypothesis claims that the parser considers an (ungrammatical) interpretation in which the quantifier takes scope over the whole sentence (making the sentence globally negative), leading to the perception that the NPI is acceptable (even with a perfectly accurate NPI-licensing mechanism). Because the error in interpretation actually occurs well before the NPI is encountered on this hypothesis, we should be able to see consequences of this interpretive error in sentences with or without an NPI. In experiment 3 (30 subjects; 36 items; data collected using Amazon Mechanical Turk) we tested whether sentences with negative quantifiers inside of relative clauses were interpreted as globally negative using comprehension question responses. While we find slight increases in globally negative interpretations for sentences with negative quantifiers and no NPI (about 8% of accepted sentences are interpreted as negative) compared to baselines (about 1%), this difference isn't enough to explain the illusion rate (about a 27 percentage point difference in acceptance rates). It also isn't nearly large enough to explain the high rate of negative interpretations of NPI illusion sentences (over 75%). Thus we see no evidence that globally negative interpretations established as a result of ungrammatical quantifier scope are responsible for illusions.

Conclusion. Across 3 experiments we establish that the profile of NPI illusions is consistent with an online licensing mechanism that relies on context meanings, making reference to scalar pragmatic alternatives. We provide evidence supporting the hypothesis that illusions arise not because of the nature of memory retrievals but because of difficulty inhibiting nearby contexts with the appropriate semantics and pragmatics.

References. [1] Chierchia (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. [2] Drenhaus, Saddy, & Frisch (2005). Processing negative polarity items: When negation comes through the backdoor. [3] Homer (2012). Domains of polarity items. [4] Israel (1996). Polarity sensitivity as lexical semantics. [5] Kadmon & Landman (1993). Any. [6] Parker & Phillips (2016). Negative polarity illusions and the format of hierarchical encodings in memory. [7] Vasishth, Brüssow, Lewis, & Drenhaus (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. [8] Xiang, Grove, & Giannakidou (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences.