RESEARCH PAPER

Real-time social reasoning: the effect of disfluency on the meaning of *some*

Jia E. Loy [ID] · Hannah Rohde · Martin Corley

Abstract The scalar quantifier *some* is locally ambiguous between pragmatic (some-but-not-all) and literal (some-and-possibly-all) meanings. Although comprehenders typically favour an eventual pragmatic interpretation, debate persists regarding what factors influence interpretation, the time course of comprehension, and whether literal meaning takes precedence. We investigate how the interpretation of *some* depends on social reasoning derived from a speaker's manner of delivery. Specifically, we test the effect of disfluency on the derivation of meaning in a context where hesitation may signal speaker embarrassment due to potential face-loss associated with the literal meaning of "some". Participants ($n = 24$) viewed displays comprising two different snack quantities while hearing a recorded utterance describing how much a speaker had eaten. Critical utterances ($n = 16$) contained the quantifier *some*, half with a filled pause disfluency ("I ate <*uh*>, some oreos"). Participants' eye and mouse movements showed (via empirical logit regressions) that fluent utterances yielded a bias toward a pragmatic interpretation, while disfluency attenuated this bias in favour of the literal meaning (where the speaker ate all the oreos). Crucially, this difference emerged rapidly post-onset of *some*. Taken together, our findings do not support a literal-first account of scalar comprehension, but rather, suggest that *some* is interpreted rapidly in a context-dependent manner.

Keywords Disfluency · Scalar implicature · Discourse context · Eye-tracking · Mouse-tracking

Present Address:
J. E. Loy (✉) · H. Rohde
Linguistics & English Language, PPLS, University of Edinburgh, 3 Charles Street,
Edinburgh EH8 9AD, Scotland, UK
e-mail: jia.loy@ed.ac.uk

M. Corley
Psychology, PPLS, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland, UK

Introduction

Successful comprehension often requires listeners to infer a speaker's meaning from an ambiguous utterance. One situation where ambiguity may arise is in the case of scalar expressions, which can elicit so-called literal or pragmatic interpretations. Take the following sentences for example:

1(a)   Some coconuts grow on trees.
(b)   Some, but not all, coconuts grow on trees.
(c)   Some, and possibly all, coconuts grow on trees.

Sentences such as (1a) typically evoke the interpretation (1b). This pragmatic inference from *some* to *not all* is taken to reflect reasoning based on the Gricean maxim of Quantity (Grice 1975): A cooperative

speaker could have made (1a) more informative by instead saying *all*; the fact that they didn't implies that they were not in a position to do so, triggering the pragmatic interpretation (1b). Nevertheless, given a lower-bound meaning of *some* to be *at least one*, a literal interpretation of (1a) would yield the meaning (1c). Ambiguity between the two meanings has yielded a long line of research investigating how listeners derive the speaker's intended meaning (e.g., Breheny et al. 2006; Carston 1998; Degen 2015; Horn 1984; Russell 2006; Van Tiel et al. 2016), and the time course with which each meaning arises during comprehension (e.g., Bott and Noveck 2004; Degen and Tanenhaus 2014; Huang and Snedeker 2011, 2009a, b; Noveck and Posada 2003; Tomlinson et al. 2013; see Chemla and Singh 2014a, b for reviews). Although adult comprehenders typically favour an eventual pragmatic interpretation (Grice 1975; Horn 1972; Noveck 2001; Papafragou and Musolino 2003; Van Tiel et al. 2016), a continuing debate centres around the time course of this interpretation and how and when context plays a role.

The timecourse of comprehension

The intuitive ease with which comprehenders deduce *some* to mean *not all* has led some researchers to propose that these pragmatic inferences are stored in the lexicon and computed automatically by a dedicated grammatical system (Chierchia 2004, 2006; Levinson 2000). These accounts predict that the pragmatic meaning arises automatically and immediately, although it may be overridden by the literal meaning later (Bott and Noveck 2004; Breheny et al. 2006). This view is also consistent with broader evidence that comprehenders can make rapid pragmatic inferences about the speaker or discourse, often from the earliest moments of comprehension (Grodner and Sedivy 2011; Hagoort et al. 2004; Hanna and Tanenhaus 2004; Kurumada et al. 2014; Loy et al. 2017; Rohde and Horton 2014; Van Berkum et al. 2008).

Evidence from other studies, however, suggests that the pragmatic interpretation of *some* is slow and effortful for listeners to access, arising after its literal counterpart is derived. In a series of eye-tracking studies, Huang and Snedeker (2009a, b, 2011) demonstrated that listeners interpret *some* as compatible with its literal meaning before converging on the pragmatic interpretation. In the *some* condition, participants told

to follow audio instructions such as "Point to the girl that has *some/two/all/three* of the ice cream sandwiches" were initially equally likely to fixate the referent compatible with a literal meaning (girl with all of the ice cream cones) and a pragmatic meaning (girl with a subset of the ice cream sandwiches). Fixations to the pragmatic target did not reliably exceed chance until 1,100 ms post-onset of *some*, suggesting that the pragmatic meaning was not available during early processing to rule out a literal interpretation. Similar processing costs for pragmatic *some* have been observed in self-paced reading and sentence verification tasks (Bott et al. 2012; Bott and Noveck 2004; Breheny et al. 2006). Bott and Noveck showed, for example, that participants took longer to evaluate ambiguous sentences such as "Some elephants are mammals" when instructed to assume a pragmatic interpretation of *some*, compared to those instructed to assume a literal interpretation. Moreover, when given no restrictions on how to interpret *some*, participants who intuitively responded with the pragmatic interpretation took longer than those who responded with the literal meaning (cf. Noveck and Posada 2003). Together, these findings demonstrate a temporal delay associated with pragmatic *some*, suggesting some form of costly pragmatic enrichment applied to an initial literal interpretation.

This position, however, has not gone unchallenged. Using a similar paradigm and methods to Huang and Snedeker, Grodner et al. (2010) showed that the pragmatic meaning of *some* can arise from the earliest stages of comprehension, with no evidence of precedence by the literal meaning. Grodner et al. hypothesised that one reason for the pragmatic delay observed by Huang and Snedeker could have been the inclusion of trials with an exact number ("Point to the girl with *two* of..."), thereby reducing the felicity of utterances on *some*-trials, where the target would also have two of a set of objects. Modifying their design to eliminate instructions with exact quantities, Grodner et al. showed that listeners' eye movements converged on the appropriate target within 200–300 ms post-quantifier onset, and were equally fast in pragmatic *some* and literal *all* conditions. Grodner et al. suggested that the pragmatic delay observed by earlier studies is not inherent to the actual generation of the pragmatic inference, but rather arises from the difficulty in integrating its meaning with available contextual information, such as considerations about

alternative forms that the speaker may have used. They argue that with appropriate and adequate contextual support, processing delays associated with the pragmatic interpretation disappear.

The role of context

Grodner et al.'s results forefront the relevance of context in the comprehension of scalar expressions. This notion of contextual-sensitivity has been similarly highlighted by a number of other researchers (Bonnefon et al. 2009; Breheny et al. 2006; Chierchia 2004; Degen and Tanenhaus 2014). Breheny et al. (2006) showed, for instance, that the speed with which *some* is comprehended is dependent on discourse context. Experiment 3 manipulated the context in which participants read sentences (in Greek) containing a scalar trigger "some of the Xs", and measured their reading times on a subsequent target segment *the rest*, referring to the complement set (the remaining Xs) evoked by a pragmatic interpretation of *some* (see Example 2).

- 2(a) Mary asked John whether he intended to host all his relatives in his tiny apartment. John replied that he intended to host *some of his relatives*.
- (b) Mary was surprised to see John cleaning his tiny apartment and she asked why. John replied that he intended to host *some of his relatives.*
- Target *The rest* would stay in a nearby hotel.

In upper-bound contexts (2a), where a reading of *at least one* is sufficient and an interpretation of *possibly all* is disfavoured by the contrast with *all*, participants' reading times were faster compared to lower-bound contexts (2b), which were compatible with a literal interpretation of *some*. This suggests that the context in which *some* occurs influences comprehension depending on whether it supports a literal or a pragmatic interpretation. In a similar vein, participants in Bergen and Grodner (2012) were slower to read sentences containing *some* in cases where they thought the speaker of the sentence would know that the stronger *all* statement was true, compared to cases where they thought the speaker might (but did not necessarily) know that the stronger statement was true. Although these results still suggest a delay associated

with the pragmatic meaning, they nevertheless highlight the role of context, in that the speed with which *some* is processed is modulated by knowledge of the discourse or the speaker's state.

Another form of context that has been found to be relevant is visual context. Degen (2015) investigated the interpretation of *some* using a "gumball paradigm", in which participants were asked to rate the naturalness of statements such as "You got *some* of the gumballs" depending on how many gumballs they saw being partitioned (out of a full set of 13) into a lower chamber of a gumball machine. Smaller sets (1–3 gumballs) as well as full sets (all 13 gumballs) received lower ratings than mid-range (5–8) sets. Degen (2015) proposed that the size of the partitioned set increased the salience of alternative quantifiers (e.g., *two, many, most, all*) for small and large sets, creating listener expectations about the use of *some* with different set sizes. Ratings also decreased when the experiment included filler sentences which used number terms for small and large sets, suggesting that participants took into account lexical alternatives which the speaker had previously used instead of *some*. This idea is also consistent with Grodner et al.'s finding that the elimination of sentences with exact quantities facilitated participants' processing of pragmatic *some*.

Bonnefon et al. (2009) demonstrate context-dependency using a different class of context: one in which the speaker's politeness goals may be relevant to a listener (Brown and Levinson 1987). They showed in an off-line task that sentences which may represent a face-threat to a listener (cf. Goffman 1967) (e.g., "Some people hated your poem"), were more likely to generate the literal interpretation that *everyone* hated your poem. In contrast, a face-boosting version of the same sentence (e.g., "Some people loved your poem") was more likely to elicit the pragmatic interpretation that *not everyone* loved your poem. Bonnefon et al. suggest that in contexts where politeness concerns may be relevant to the discourse, individuals may construe *some* as a politeness device employed by the speaker to mitigate the effects of a potential face-threat, leading listeners to consider the more face-threatening interpretation of the utterance.

Manner of delivery as a contextual cue

While existing literature provides evidence that scalar comprehension may depend on its global context of

occurrence, comparatively little work has examined how interpretation may be affected by a more local source of variation—that of the speaker's manner of delivery. There are reasons to believe that manner of delivery may influence the interpretation of *some* due to a listener's reasoning about the social context. Research outside of scalar comprehension highlights the perceptual relevance of spoken manner on various aspects of pragmatic comprehension (e.g., Kurumada et al. 2014; Van Berkum et al. 2008). A particular focus of recent research has been on disfluencies, or pauses in the fluent flow of an utterance, whether silent or filled with a sound such as *um* or *uh*.

Disfluencies tend to be produced during more complex dialogue turns (Bortfeld et al. 2001), before longer utterances (Oviatt 1995), and at major discourse boundaries (Boomer 1965). Speech task complexity is often cited as an important predictor of filled pauses (e.g., Barr 2001; Merlo and Mansur 2004; Womack et al. 2012). Filled pauses are used equally often across demographic groups, compared to filler words such as *like* or *you know* (Laserna et al. 2014), suggesting that they may be automatically produced as a consequence of cognitive load. However, cognitive load doesn't appear to be the sole predictor of disfluency. For example, Smith and Clark (1993) noted that an *um* can precede the single-word answer to a question; this may be providing a speaker with time to retrieve the answer to a question while simultaneously cueing the listener to tune in (Clark and Fox Tree 2002).

Following an utterance-medial disfluency, listeners are more likely to predict a mention of an object that is new to the discourse (Arnold et al. 2004) or requires a longer noun phrase to name (Arnold et al. 2007). The N400 ERP response, taken to index the consequences of processing something unpredictable in context, is attenuated when an unpredictable noun directly follows a filled (Corley et al. 2007) or silent pause (MacGregor et al. 2010).

As well as their effects on prediction, disfluencies such as *um* or *uh* have been found to influence listeners' pragmatic inferences about whether or not a speaker is lying (see Zuckerman et al. 1981, for an early review). Importantly, recent findings show that listeners very quickly integrate such paralinguistic cues from the earliest moments of comprehension to shape their overall utterance interpretation (King et al. 2017; Loy et al. 2017), highlighting the speed with

which manner of delivery can affect meaning construction via a process of social reasoning.

Preliminary work by Bonnefon et al. (2015) also suggests that manner of delivery may influence comprehenders' *eventual* interpretations of *some* within a relevant social context. Based on earlier findings that a speaker's politeness goals affect the meaning of *some* for comprehenders (Bonnefon et al. 2009), Bonnefon et al. (2015) hypothesised that silent pauses might influence listeners' interpretations of the quantifier within such a context, by functioning as a social cue to shift expectations toward unpleasant information.

(3)   Yesterday, you pitched an idea to a group of five persons. Today, you ask Bob (who was in the group) what people thought of your idea. Bob *<stays silent for a few seconds. Then he>* replies: "Some people hated your idea."

The study manipulated the description of whether or not a speaker remained silent in a scenario before delivering a face-threatening expression [see example (2)], and asked participants to rate the extent to which the statement warranted a literal interpretation (i.e. possibly everyone hated your poem). Scenarios in which the speaker was described as remaining silent before speaking received higher ratings in favour of the more unpleasant interpretation—in this case the literal interpretation of *some*. Conversely, with a face-boosting expression (e.g., "Some people loved your idea"), the same pause description yielded higher ratings in favour of pragmatic interpretation of *some* (i.e. not everyone loved your idea).

Bonnefon et al.'s (2015) results are indicative on two fronts. Firstly, in line with previous work, they outline a relationship between context and the comprehension of scalar expressions (Bonnefon et al. 2011, 2009; Breheny et al. 2006; Cummins and Rohde 2015; Degen and Tanenhaus 2014; Feeney and Bonnefon 2012; Katsos and Bishop 2011). Second, they provide prima facie evidence that given a relevant context, interpretations of *some* may vary with the manner in which the utterance is presented. However, Bonnefon et al. used a task in which participants were explicitly asked to consider the possibility of the literal meaning of *some*, following a pause which could be presumed to be relevant to the interpretation of the utterance, given that it was explicitly described. As

such, the results likely reflect metalinguistic reasoning about a manner of delivery to which participants' attention had been drawn. While these findings establish a relationship between the manner of delivery of an expression and a comprehender's eventual considered interpretation of *some*, they leave open the question of whether such cues influence the interpretation of more naturally-produced utterances during real-time comprehension.

The present study

The present study investigates whether spoken manner of delivery influences a listener's interpretation of the ambiguous quantifier *some*, during the moment-to-moment processing of the linguistic expression. In the experiment, listeners make an implicit choice between a literal and a pragmatic interpretation of *some* based on a speaker's fluent or disfluent delivery of the scalar expression. In a similar manner to Bonnefon et al. (2015), we established a context that exploited the concept of face (Goffman 1967)—in this case, one in which a literal interpretation of *some* would threaten the positive self-image of the speaker. To achieve this, we invented a cover story about a fictitious experiment investigating greed and snacking habits.

The cover story was as follows: We described a set of participants who were provided with a variety of snacks to eat while watching a documentary film. They received no instruction other than that they could eat as much or as little as they liked, and had to answer questions about the film in a verbal interview afterwards. We described that the study's motivation of investigating greed was revealed after the documentary, and then the fictitious participants were asked to report how much of each snack they had eaten (e.g., "I ate five oreos").

Participants in the current experiment were told that they would hear recordings of people who had taken part in the earlier experiment. This set up a context in which speakers who had consumed all of a snack might plausibly exploit the ambiguity of *some* to avoid face-loss through an admission to greed. Crucially, speakers might be disfluent as a by-product of the calculation of the potential threat to their positive self-image. In other words, "I ate *uh*, some oreos" could be taken to mean that a speaker ate *all* of the oreos but is embarrassed to admit it.

Each recorded utterance was played while participants viewed a visual display comprising two plates,

with each plate depicting a quantity of one of the snack items. Participants were tasked with clicking on the plate that depicted what was left behind, based on the speaker's description. We measured their eye- and mouse-movements during each trial. Critical utterances made use of the quantifier *some*. Half of these included a filled pause disfluency, chosen to avoid any likelihood that, once items were being performed rather than described, a silent hesitation could be construed as a prosodic pause (e.g., Ferreira 2007).

We tested participants' interpretations of *some* using an Ambiguous display, where both plates were compatible with the utterance— an empty plate (no snacks remaining) was compatible with the literal interpretation, while a second plate with a number of snacks remaining was compatible with the pragmatic meaning. Based on existing research, we expected an overall bias toward a pragmatic interpretation of *some* (Noveck 2001; Papafragou and Musolino 2003; Van Tiel et al. 2016). Importantly, on the basis that the face-saving context would induce listeners to interpret a disfluency as a signal that the speaker was avoiding face-loss, we expected filled pauses to yield a higher rate of the literal interpretation, and thus an increase in fixations on, mouse movements towards, and mouse clicks on the empty plate.

A potential concern with our predictions is that disfluency may affect the interpretation of *some* for other reasons. For example, previous work has shown that listeners can interpret disfluency as a signal of simple deception, where the speaker means the opposite of what they say (Akehurst et al. 1996; Zuckerman et al. 1981). These effects have been shown to rapidly influence comprehension, emerging almost as soon as a listener can infer meaning based on the unfolding linguistic input (Loy et al. 2017). Thus, given the typical inference that *some* means not-all, a disfluency construed as a signal of deception might quickly bias listeners toward the plate compatible with the atypical literal meaning, in this case the empty plate. In an attempt to rule out alternative accounts of our findings, we included a second condition in the experiment, which included a Control display. In this condition, one plate had a few snacks remaining, as in the Ambiguous condition (for ease of reference, we refer to this plate throughout as the 'pragmatic' plate). However the second plate—a distractor—always had 6 snacks remaining, corresponding to one piece having been eaten (see Fig. 1). We reasoned that if listeners

interpret disfluency within the social context established by the experiment, disfluency should be associated with face-saving. This should lead to a bias towards the plate with the most snacks missing following disfluency. Crucially, any bias towards the distractor plate following disfluency in the Control condition would suggest that a face-saving account could not be sustained; disfluency would only be associated with the removal of *a single* snack under different circumstances, such as those in which disfluency signals deception.

Participants' eye- and mouse-movements were recorded on each trial, as well as their eventual interpretations (plate clicked) and response times. We expected disfluency to result in more movements towards, and clicks on, the distractor (empty, literal interpretation-compatible) plate in Ambiguous trials. In Control trials, we did not expect an increase in movements towards or clicks on the distractor (one snack removed, incompatible with face-saving) following disfluency. Of particular interest were the timings with which any biases in movements emerged: The eye- and mouse-tracking records allow us to establish when, relative to encountering the ambiguous *some*, listeners begin to commit to a particular interpretation.

Method

Participants

Twenty-four self-reported native British English speakers took part in the experiment. Sample size was based on those of Loy et al. (2017, $n = 21, 22$), in which two experiments included eye- and mouse-movement analyses comparable to those of the present design. All participants were right-handed mouse users with normal or corrected-to-normal vision. An additional 3 participants were tested, but their data were not included in our analyses because they suspected the authenticity of the cover story (2) or that the audio had been scripted for the experiment (1; determined during debrief). Participants were recruited from the University of Edinburgh community and each received £5 for participation. All participants provided informed consent in accordance with the university's Psychology Research Ethics Committee guidelines (ref no.: 136-1617/1).

Materials and design

Eight different types of snacks were used as referents in the experiment. The cover story established a starting quantity of 7 for each snack item (see Fig. 2). The 8 snacks were chosen based on a pre-test of 12 snacks, in which respondents indicated the likelihood that they would eat up to 7 pieces of each snack in one sitting.

On each trial, participants saw a visual display comprising two plates, each depicting a quantity (range 0–7) of one of the snack items. This quantity represented the number of pieces of the snack that remained (out of 7). The name of the snack was displayed below each plate to avoid ambiguity in cases where 0 pieces remained. Each display was accompanied by a recording of a speaker describing how much of a snack they had eaten. The utterances were produced by 8 speakers (4 male; all native British speakers), each contributing 8 utterances (one per snack), for a total of 64 utterances used in the experiment. Two out of each speaker's 8 utterances were critical utterances; the other 6 were fillers. Snacks were balanced across speakers such that each snack only occurred as a referent in two critical

Fig. 1 Example of displays used in **a** ambiguous and **b** control trials. In both displays the plate on the left depicts the pragmatic interpretation of *some*
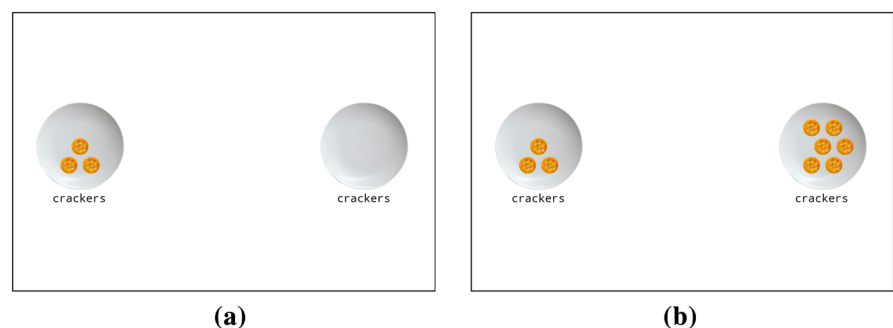
(a)   **(b)**

Fig. 2 Snack items used as referents in the experiment

Each participant was given 7 pieces of each of the following snack items:

brownie bites   crackers   mini flapjacks   gummy fish

jelly babies   marshmallows   oreos   mini pretzels

utterances, each by a different speaker, with no two speakers associated with the same two critical referents.

Speakers were recorded individually using a Zoom H4N digital recorder. For the first recording, a female speaker who had prior experience producing disfluent speech materials read the sentences from a script. All subsequent speakers were recorded via a shadowing procedure in which they listened to the first speaker's recordings and imitated the speech, utterance by utterance (cf. Bosker et al. 2014; Hanulíková et al. 2012). From each speaker's recordings, a filled pause disfluency ("uh") from the disfluent utterance that sounded the most natural was excised and cross-spliced into each fluent critical utterance to create a disfluent counterpart. This ensured that each speaker's critical utterances were identical (bar disfluency manipulation) across the fluent and disfluent conditions. All utterances were normalised to have the same mean acoustic intensity.

On critical utterances, the speaker used *some* to describe how much of the snack they had eaten. In Ambiguous displays, this was compatible with two interpretations of *some*—a pragmatic interpretation depicting 2, 3 or 4 remaining pieces of the referent (i.e. corresponding to 5, 4 or 3 pieces having been eaten), and a literal interpretation depicting 0 pieces. These quantities for the pragmatic interpretation were chosen based on evidence that *some* is perceived as most natural when used to reference intermediate set

sizes (e.g., 6–8 out of 13 gumballs; Degen 2015). In Control displays, the first plate depicted 2, 3 or 4 remaining pieces, as in Ambiguous displays. The second plate, a distractor, contained 6 pieces of the referent to illustrate one piece having been eaten—an interpretation intended to be incompatible with any bias to interpret *some* literally in a face-saving context. Half of the utterances accompanying each display were fluent ("I ate some crackers") and the other half disfluent ("I ate *uh,* some crackers"). Hence, the study followed a 2 (manner: fluent/disfluent) × 2 (display: Ambiguous/Control) within-subjects design, with critical utterances counterbalanced across 4 lists. The quantity displayed on the pragmatic plate on each trial was chosen at random from a list, with 2, 3 and 4 represented equally across conditions. Within each condition, the pragmatic plate appeared on the left and the right an equal number of times.

The 16 critical trials were randomly presented together with 48 fillers. To increase variability, these included a number of manipulations in the speaker's manner. Half of the filler utterances were fluent; the other half contained some other form of disfluency (e.g., a prolongation: "I *aate...* four oreos") or a hedge suggesting uncertainty about the exact quantity eaten (e.g., "I ate *maybe...* three jelly babies"). Filler utterance types were distributed across speakers such that each speaker produced an even mix of fluent and non-fluent filler utterances. Filler trials also varied at the level of display. In half the filler displays, the

distractor plate depicted a different quantity of the referent snack. In the other half, the distractor plate depicted a different snack, with half of these depicting the same quantities of each snack. This manipulation had the purpose of discouraging listeners from focusing only on the quantifier to disambiguate between the two plates on each trial. Filler displays were distributed such that each of the 8 speaker's filler utterances were accompanied by a variety of filler display types. The same set of filler trials was used in all 4 experimental lists.

Procedure

The experiment was presented using OpenSesame 3.1.0 (Mathôt et al. 2012) on a 21 in. CRT monitor. Eye movements were monitored using an Eyelink 1000 Tower Mount system sampling at 500 Hz. Mouse coordinates were sampled at 50 Hz.

Participants were first briefed on the cover story which established the context in which the utterances were produced. To corroborate the story, the instructions included a photo ostensibly taken of a participant taking part in the fictitious experiment.

Following the instructions, the eyetracker was calibrated. Between trials, participants underwent a manual drift correction using a central grey fixation dot. After this, the dot turned red for 500 ms to signal the start of the trial. Each trial began with a 1000 ms presentation of the two full plates containing 7 items each. This served to remind participants of the starting quantity of each snack item. The two plates were centred vertically and positioned horizontally left and right on the screen. This was followed by a 1000 ms preview of the actual quantities associated with each snack for the trial. After this, a mouse pointer appeared at the centre of the screen and playback of the utterance began. Participants were instructed to click on the plate depicting the quantity remaining based on the speaker's description of what they ate. For example, if the participant heard "I ate five oreos", they would click on the plate depicting two oreos. There was no feedback except in cases where participants failed to click on a plate within 5000 ms post-utterance offset, following which they received a message to respond more quickly. Participants underwent 4 practice trials and were given the opportunity to ask questions afterwards, before the main experiment

began. None of the practice trials included the word *some*.

After the experiment, participants completed a post-test questionnaire in which they were asked (a) whether they noticed anything striking about the audio or visual stimuli, and (b) what they believed the experiment was investigating. Any participant who answered "yes" to the first question was asked to elaborate verbally on this during the debrief; a note was made if they mentioned being suspicious that the disfluencies were not naturally produced or that the audio had been scripted for the experiment. Participants were also questioned during debrief on whether they had suspected the authenticity of the cover story after the experimental manipulation had been revealed. Data from participants who questioned the authenticity of the recordings (1 participant) or did not believe the cover story (2) were excluded from analysis.

Results

Statistical analyses were carried out in R Version 3.3.3 (R Core Team 2017). Our analyses focused on listeners' final interpretations of *some* for each utterance (plate clicked), response times, eye movements and mouse movements. For each dependent variable, we modelled the effect of manner of delivery (fluent/disfluent) individually for each display type (Ambiguous/Control). To evaluate the difference in the effect of manner on the two display types, we also ran an interaction model taking into account both manner and display as fixed effects. Predictors were mean centred in all analyses.

Logistic regression was used to model the binary outcome of which plate participants clicked on. The distribution of responses reflected an overwhelming bias toward a pragmatic interpretation of *some*. To avoid spurious ceiling effects, a generalised linear model by robust methods was fit using the glmrob function from the robustbase package (Maechler et al. 2016). This approach produces more robust estimation of regression parameters in cases where inference based on maximum likelihood may yield unreliable results (Cantoni and Ronchetti 2001). Linear mixed effects regression was used to model participants' response times, using the lmer function from the lme4 package (Bates et al. 2014). Models included by-

Table 1 Breakdown of mouse clicks (raw count) recorded on each plate and mean response times (in ms) following fluent/disfluent utterances on Ambiguous/Control displays

	Plate clicked		Response time
	Pragmatic	Competitor[a]	
Ambiguous			
Fluent	94	2	1842 (74.6)
Disfluent	86	10	2103 (73.0)
Control			
Fluent	95	1	1914 (66.0)
Disfluent	94	2	1807 (63.6)

Standard errors are in parentheses

[a] On Ambiguous displays the competitor plate represented the literal meaning of *some*; on Control displays the competitor was a distractor incompatible with any meaning of *some*

subjects and by-items random intercepts and slopes for manner and display.

Eye-tracking records were averaged into 20 ms bins, each comprising 10 samples, prior to analysis. Data were coded in terms of fixations toward either one of the plates or areas outside of both. The proportion of fixations to each plate out of the total sum of fixations was computed for each time bin. Mouse-tracking analysis only took into account the X coordinates. For each sample, the distance travelled by the mouse was computed by taking the absolute difference between the X coordinates of the current and previous samples. The data were coded for direction of movement toward either one of the plates for each bin, and the cumulative distance participants had moved the mouse toward each plate was computed by summing over the distance travelled in each direction up until that time bin (taking into account all previous mouse movements in that direction on that trial). For each plate, we then calculated a proportion-of-movement measure, defined as the distance travelled by the mouse pointer towards the given object, divided by the total distance travelled (regardless of X direction).

To evaluate whether manner of delivery influences listeners' processing of *some* during real-time comprehension, eye- and mouse-tracking data were analysed over an 800 ms time window beginning from 200 ms post-quantifier onset. This window corresponds to the duration of the quantifier and subsequent referent, taking into account the 200 ms it typically takes to program and execute an eye movement (Matin

et al. 1993), and ending just before the average utterance offset (1071 ms). Models for this window were fitted using empirical logit regression (Barr 2008), taking as the dependent variable the difference between the e-logit of fixations (or mouse movements) to the two plates on each trial. Fixed effects included time, manner and display (all predictors mean centred). All models included by-subjects and by-items random intercepts and slopes for all predictors.

Click responses

Table 1 shows the breakdown of mouse clicks recorded on each plate following fluent and disfluent utterances on each display. The last column shows the mean response time (in ms) measured from the onset of *some*.

For Ambiguous displays, a robust logistic regression on the outcome of mouse clicks showed an effect of manner of delivery. Disfluent utterances resulted in fewer clicks on the pragmatic plate (and therefore more clicks on the literal plate), $\beta = -1.70$, $SE = 0.86$, $p = .049$. A linear mixed effects regression on listeners' response times showed an effect of manner of delivery. Listeners were slower to click on a plate following a disfluent utterance, $\beta = 260.46$, $SE = 74.81$, $t = 3.48$. For Control displays, there was no effect of manner on listeners' mouse clicks ($p = .6$) or response times ($t = -1.38$). These results provide no evidence to support the disfluency-signals-simple-deception hypothesis.

A robust logistic regression on listeners' mouse clicks including both manner and display as predictors showed no effect of either, nor any interaction between the two (all $p > .1$). A linear mixed effects regression on response times showed no main effects of manner or display, but yielded a manner by display interaction, $\beta = 368.15$, $SE = 106.78$, $t = 3.45$, reflecting the longer time taken by listeners to click on a plate following a disfluent utterance on Ambiguous displays.

Eye movements

Figure 3 shows the proportion of fixations to each plate over time until 2000 ms post-*some* onset, by which point participants had typically moved the mouse over one of the two plates. The pattern of fixations on Ambiguous displays demonstrates a

Fig. 3 Proportion of fixations to each display plate over time during fluent and disfluent utterances for Ambiguous (top) and Control (bottom) displays. Shaded areas represent ±1 SE of the mean. On Ambiguous displays the competitor represented a literal meaning of *some*; on Control displays the competitor was a distractor incompatible with any meaning of *some*
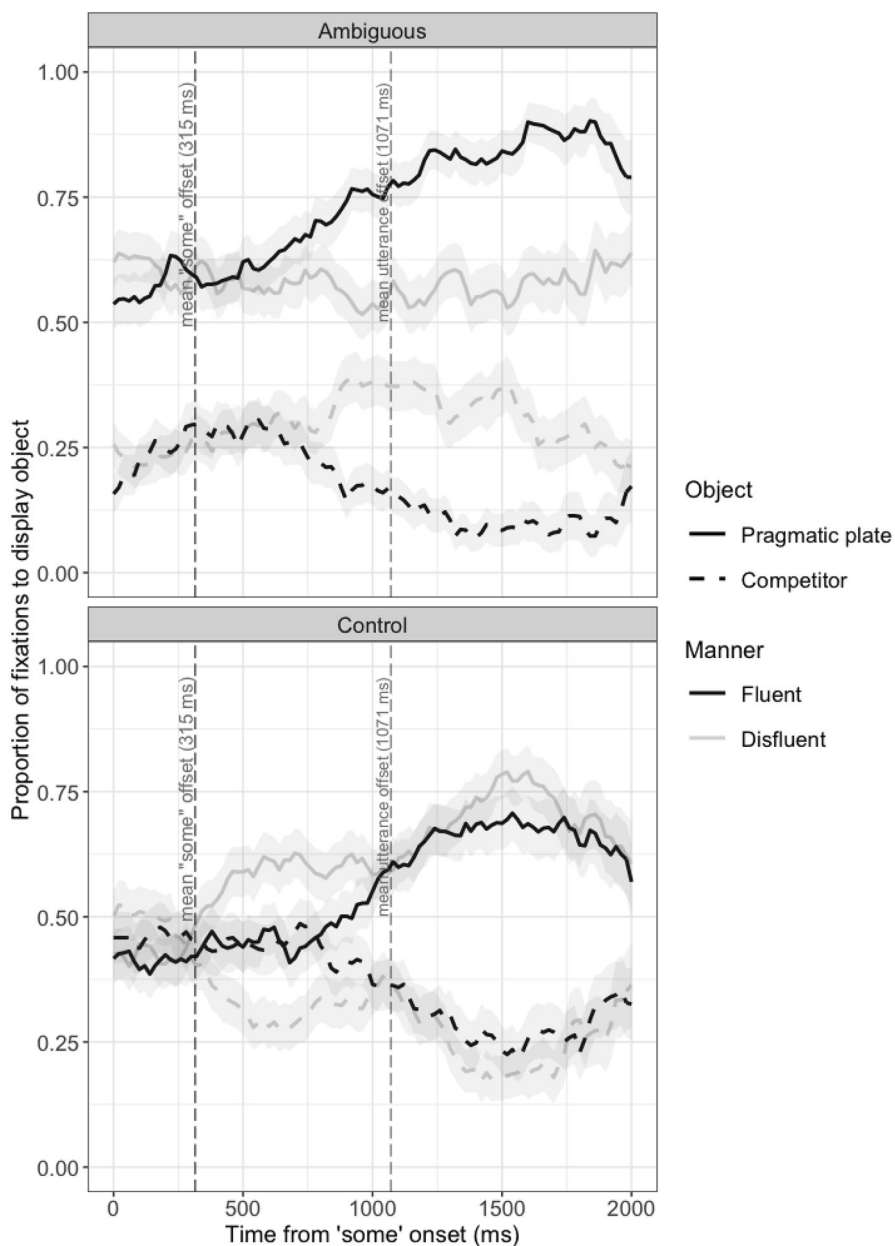
baseline bias toward the pragmatic plate relative to the literal plate. This likely reflects a preference to look at plates with objects over empty plates, and is consistent with earlier studies which report a fixation bias to the image with the largest quantity of items prior to disambiguation (Grodner et al. 2010; Huang and Snedeker 2009b). As predicted under a model in which disfluency is interpreted in a social context, there is an influence of manner of delivery. Fluent utterances led to a rapid rise in fixations to the

pragmatic plate after the onset of *some*; on disfluent utterances, this increase was attenuated. This difference was reflected in a time by manner interaction, $\beta = 3.55$, $SE = 0.46$, $t = 7.71$.

In contrast, on Control displays, disfluent utterances saw an earlier rise in fixations to the pragmatic plate compared to fluent utterances, as evidenced by a time by manner interaction, $\beta = -1.19$, $SE = 0.50$, $t = -2.39$. The difference in the effect of manner on the Ambiguous and Control displays was confirmed by a

three-way time by manner by display interaction, $\beta = 4.78$, $SE = 0.68$, $t = 7.09$. We note that this result nevertheless does not support a disfluency-signals-simple-deception hypothesis, which predicts a fixation bias to the competitor plate following disfluent utterances. We return to this effect in the Discussion.

Mouse movements

Figure 4 shows the proportion of mouse movements (in terms of distance travelled) toward each plate over time until 2000 ms post-*some* onset. Mouse movements follow a pattern compatible with the fixation data. On Ambiguous displays, participants' mouse movements exhibited a preference for the pragmatic plate over the literal plate following fluent utterances, which was attenuated during disfluent utterances, $\beta = 2.87$, $SE = 0.34$, $t = 8.44$. In contrast, on Control displays mouse movements were characterised by a greater preference for the pragmatic plate over the competitor during disfluent utterances, $\beta = -1.41$, $SE = 0.31$, $t = -4.57$. This effect aligns with the early fixation bias to the pragmatic plate following disfluent utterances on Control displays. As with the eye movements, the difference in the effect of manner on listeners' mouse movements during the two displays was confirmed by a three-way time by manner by display interaction, $\beta = 4.22$, $SE = 0.49$, $t = 8.63$.

Discussion

This study set out to test whether listeners' interpretations of the ambiguous quantifier *some* vary with the speaker's manner of delivery. Like Bonnefon et al. (2015), we made use of a social context that exploited the concept of face—in this case one where snacking is associated with greed, which in turn threatens the positive self-image of a speaker. This allowed us to establish a context in which a speaker's disfluency could be perceived as a social cue that signals a potential face-loss for the speaker. Our results suggest that listeners did indeed assign this social meaning to speakers' disfluencies. Fluent utterances yielded an overwhelming bias toward the pragmatic interpretation. This pattern follows a robust trend established in the literature for adult listeners to assign to *some* a meaning of *not all*. However, when the literal

meaning—the plate associated with the socially dispreferred meaning of having greedily eaten all the snacks—was available as an alternative interpretation, disfluency attenuated the bias toward the pragmatic interpretation. This was apparent in Ambiguous displays, where disfluent utterances led to a decrease in the proportion of mouse clicks on the pragmatic plate in favour of the literal plate, as well as a shift in eye- and mouse-movements in the same direction.

Under an alternative account for the effect of manner, such as one of simple deception, disfluent utterances in the Control condition should have elicited a bias toward the competitor plate, which only had one snack missing. However, we found no evidence of such a bias. Instead, listeners' mouse clicks on the pragmatic plate were at ceiling for both fluent and disfluent utterances in the Control condition, while their eye and mouse movements suggest that disfluent utterances in fact led to an earlier bias toward the pragmatic plate compared to fluent utterances. This is consistent with the view that disfluency is associated with a potential face-loss for the speaker. On this view, the disfluency signals in the fictitious context that the speaker has done 'something bad'; as more linguistic information becomes available, this is combined with the conceptual and visible context to form a coherent interpretation of the utterance, in which the implication is that the speaker has eaten more snacks (pragmatic plate) than one (distractor).

Obviously, the present findings are limited to scalar *some*, which has been an important testbed for the processing of implicatures. In principle, though, there is nothing special about the effects that we show: They should extend to other cases where a listener can infer pragmatic enrichments of the words uttered by the speaker based on the context in which the utterance was made, and the manner in which it was delivered. For example, it is easy to imagine a face-saving interpretation of an utterance such as "your poem was, uh, good".

Our results are significant on two fronts. Firstly, in line with recent work, we demonstrate listeners' sensitivity to manner-based cues such as a disfluency in shaping their on-line pragmatic hypotheses about a speaker's message. Extending previous studies which have focussed on listeners' global pragmatic inferences such as whether or not a speaker was lying (King et al. 2017; Loy et al. 2017), here we show that disfluency influences a listener's real-time

Fig. 4 Proportion of mouse movements to each display plate over time during fluent and disfluent utterances for Ambiguous (top) and Control (bottom) displays. Shaded areas represent ±1 SE of the mean. On Ambiguous displays the competitor represented a literal meaning of *some*; on Control displays the competitor was a distractor incompatible with any meaning of *some*
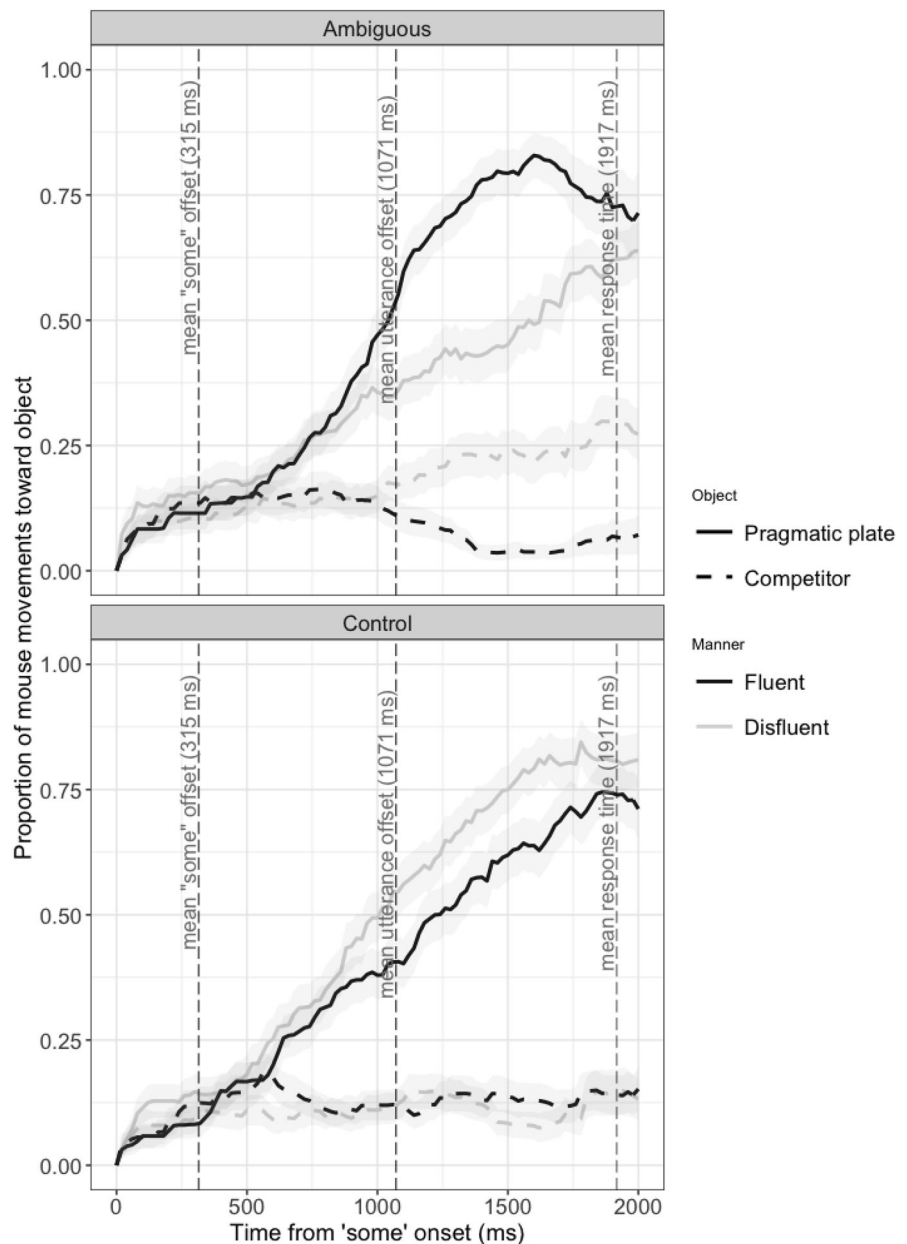
interpretation of a more local source of inferencing: meaning associated with the ambiguous quantifier *some*. These results build on earlier findings that demonstrate that listeners make rapid use of a speaker's disfluencies to evaluate syntactic ambiguity (Bailey and Ferreira 2007) or to predict semantic content (Arnold et al. 2007, 2004; Barr and Seyfeddinipur 2010; Corley et al. 2007) in an utterance, by showing an early tendency to move from a pragmatic to a literal interpretation of *some* in the face of

disfluency. Our study therefore further highlights the flexibility of the comprehension system in using manner of delivery as a cue to facilitate understanding, by drawing on different processes depending on the comprehension goals of the listener.

Secondly, and importantly, the time course of our effects demonstrates that listeners' pragmatic hypotheses about a speaker's utterance unfold during the initial stages of comprehension. The presence of disfluency attenuated responses compatible with a

pragmatic inference almost as soon as listeners could assign a meaning to *some*, and prior to the speaker's completion of the utterance. The present experiment therefore provides no evidence to support a temporal precedence of literal comprehension, in either the eye or mouse movement measures. Rather, our results suggest that listeners very rapidly take into account both manner of delivery and social context to assign meaning, via a process of reasoning about the speaker's underlying motivations (e.g., to avoid face-loss from admitting to greed).

From a methodological perspective our results are relevant to psycholinguistic research investigating the time course of language processing. Building on studies that have used mouse-tracking to replicate existing eye-tracking paradigms (e.g. Farmer et al. 2007; Spivey et al. 2005), we demonstrate that the two methods can be successfully combined in a visual world paradigm to yield a corroborating account of real-time language comprehension. In particular, the present experiment provides complementary evidence to our earlier work (King et al. 2017; Loy et al. 2017) which employed a similar paradigm to explore how disfluency modulates listeners' on-line hypotheses about a speaker's truthfulness. Thus, we show that this method can be used across various contexts to explore different pragmatic phenomena. This opens up possibilities for visual world research to employ a mouse-tracking only methodology, such as to study populations which may present challenges in eye-tracking (e.g., certain clinical or developmental groups, cf. Sasson and Elison 2012), or to obtain time course data on a large-scale through web-based data collection.

Within the field of scalar research, our findings are consistent with the view that the interpretation of *some* depends on its context of occurrence (Bonnefon et al. 2015, 2009; Breheny et al. 2006; Cummins and Rohde 2015; Degen and Tanenhaus 2014; Grodner et al. 2010). The majority of these studies have focussed on how context matters in listeners' off-line interpretation of scalar expressions; however, the present results highlight the role of context from the earliest stages of comprehension. Exploring the range of context-driven effects, such as by taking into account different types of context and how they play out during on-line comprehension, would be a useful avenue for future research on scalars.

One question arising from the current study, for example, is the nature of the interplay between social context and manner of delivery. Would the same results be observed in a context where a filled pause may serve as a different type of collateral signal (e.g., speaker uncertainty—see Brennan and Williams 1995), or in cases where its pragmatic relevance was eliminated altogether? While these questions lie beyond the scope of the current study, we propose that social and visual context deserves better attention in future research addressing the pragmatic understanding of spoken language. Our present results highlight that listeners' on-line pragmatic hypotheses regarding the meaning of *some* are modulated by the manner in which an utterance is conveyed and the context in which it is uttered. Crucially, this inference unfolds during the earliest stages of processing. Thus, we find no evidence for an "early stage" of understanding independent of context.

Compliance with ethical standards

Conflicts of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, *10*(6), 461–471.

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*(9), 578–582.

Bailey, K. G., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.),

Eye movements: A window on mind and brain (pp. 485–500). Elsevier: Oxford.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.), *Oralité and gestualité: Interactions et comportements multimodaux dans la communication* (pp. 597–600). Paris: L'Harmattan.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441–455.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models usingEigen and S4 [Computer software manual]. http://CRAN.R-project.org/package=lme4 (R package version 1.1-6)

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1450–1460.

Bonnefon, J. F., Dahl, E., & Holtgraves, T. M. (2015). Some but not all dispreferred turn markers help to interpret scalar terms in polite contexts. *Thinking & Reasoning*, *21*(2), 230–249.

Bonnefon, J. F., De Neys, W., & Feeney, A. (2011). Processing Scalar Inferences in Face-Threatening Contexts. In L. Carlson, C. Hölscher, & T. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 3389–3394)*. Austin, TX: Cognitive Science Society.

Bonnefon, J. F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, *112*(2), 249–258.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*, 148–158.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*, 123–147.

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). Native "'um"s elicit prediction of low-frequency referents, but non-native "um"s do not. *Journal of Memory and Language*, *75*, 104–116.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142.

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

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383–398.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge: Cambridge University Press.

Cantoni, E., & Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, *96*(455), 1022–1030.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carstonand & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Amsterdam: John Benjamins.

Chemla, E., & Singh, R. (2014a). Remarks on the experimental turn in the study of scalar implicature. Part I. *Language and Linguistics Compass*, *8*(9), 373–386.

Chemla, E., & Singh, R. (2014b). Remarks on the experimental turn in the study of scalar implicature. Part II. *Language and Linguistics Compass*, *8*(9), 387–399.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (Vol. 3, pp. 39–103). Oxford: Oxford University Press.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*(4), 535–590.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668.

Cummins, C., & Rohde, H. (2015). Evoking context with contrastive stress: Effects on pragmatic enrichment. *Frontiers in Psychology*, *6*(1779), 1–11.

Degen, J. (2015). Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*(11), 1–55.

Degen, J., & Tanenhaus, M. K. (2014). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667–710.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, *31*(5), 889–909.

Feeney, A., & Bonnefon, J. F. (2012). Politeness and honesty contribute additively to the interpretation of scalar expressions. *Journal of Language and Social Psychology*, *32*(2), 181–190.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22*(8), 1151–1177.

Goffman, E. (1967). *Interaction ritual: Essays in face to face behavior*. Garden City, NY: Anchor Books.

Grice, H. P. (1975). Logic and conversation. In P. Colemanand & J. L. Morgan (Eds.), *Syntax and semantics volume 3* (pp. 41–58). New York: Academic Press.

Grodner, D., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). 'Some', and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (pp. 239–272). Cambridge, MA: MIT Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115.

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Horn, L. R. (1972). On the semantic properties of logical operators in english (Unpublished doctoral dissertation) Los Angeles: University of California.

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. *Meaning, Form, and Use in Context: Linguistic Applications*, *11*, 42.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415.

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*(8), 1161–1172.

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

King, J. P., Loy, J. E., & Corley, M. (2017). Contextual effects on online pragmatic inferences of deception. *Discourse Processes*. https://doi.org/10.1080/0163853X.2017.1330041.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335–342.

Laserna, C. M., Seih, Y. T., & Pennebaker, J. W. (2014). *Um...Who like says You Know*: Filler word use as a function of age, gender, and personality. *Journal of Language and Social Psychology*, *33*(3), 328–338.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Loy, J. E., Rohde, H., & Corley, M. (2017). Effects of disfluency in online interpretation of deception. *Cognitive Science*,. https://doi.org/10.1111/cogs.12378.

Mac Gregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, *48*(14), 3982–3992.

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., & Anna di Palma, M. (2016). robustbase:Basic Robust Statistics [Computer software manual]. http://robustbase.r-forge.r-project.org/ (R package version 0.92-7)

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380.

Merlo, S., & Mansur, L. L. (2004). Descriptive discourse: topic familiarity and disfluencies. *Journal of Communication Disorders*, *37*, 489–503.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, *9*, 19–35.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*(3), 253–282.

R Core Team. (2017). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. https://www.R-project.org/

Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition*, *133*(3), 667–691.

Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of Semantics*, *23*(4), 361–382.

Sasson, N. J., & Elison, J. T. (2012). Eye tracking young children with autism. *JoVE (Journal of Visualized Experiments)*, *61*, e3675–e3675.

Smith, V., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25–38.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, *69*(1), 18–35.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, *33*(1), 137–175.

Womack, K., McCoy, W., Alm, C. O., Calvelli, C., Pelz, J. B., Shi, P., & Haake, A. (2012). Disfluencies as extra-propositional indicators of cognitive processing. In *Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics, (pp. 1–9)*.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and Nonverbal Communication of Deception. In: Advances in experimental social psychology (Vol. 14, pp. 1–59). Elsevier.

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, *6*(1), 105–114.