Contextual effects on online pragmatic inferences.

JK; JL; MC

August 2, 2016

1 Background.

intro, disfluency rates During conversation, rarely is an utterance or its delivery pre-planned. Instead, everyday speech is for the most part spontaneous and thus often disfluent, containing pauses, "um"s, "uh"s, repetitions, revisions, and mispronunciations. Naturally occurring speech has a rate of approximately 6 to 10 disfluencies per 100 words [Bortfeld et al., 2001; Fox Tree, 1995]. The disfluent nature of speech is just one of many variable aspects of *how* an utterance might be presented to a listener. Paralinguistic features of language are not merely incidental, and often an utterance's manner of delivery is apposite to the message being conveyed, from the prosodic contour [Fernald and Mazzie, 1991], to an accompanying gesture [Alibali et al., 2001]. Speaker-disfluency generally occurs as a result of an increase in cognitive load of the speaker [Bortfeld et al., 2001; **?**] and, as such, may provide information relevant to the interpretation of the message. Disfluency rates increase when utterance planning involves low-frequency words [Beattie, 1979], less-preferred syntactic structures [Cook et al., 2009], discourse new expressions [Arnold et al., 2003], and more choice of expressive alternatives [Schachter et al., 1991].

listeners sensitivity Although they are unable to accurately locate filled pauses ("um"s and "uh"s) in sentences they have just heard [Lickley and Bard, 1998], listeners' have been shown to be subtly sensitive to disfluencies [Ferreira and Bailey, 2004; Arnold et al., 2003; Barr, 2001; Corley and Hartsuiker, 2011]. Similar to their occurrences in speech production, disfluencies influence listeners' expectations about the unfolding of an utterance, affecting both the syntactic parsing of an utterance [Ferreira and Bailey, 2004; Lau and Ferreira, 2005] as well as predictions about semantic content [Barr, 2001; Arnold et al., 2003, 2004]. Listeners appear to associate speaker-disfluency with aspects of language which are perceived as more difficult to plan, displaying biases towards less predictable continuations [Corley et al., 2007]. Arnold et al. [2004] showed that when presented with a visual display, a disfluent utterance of "Click on thee, uh ..." resulted in listeners fixating more upon discourse-new objects. In Arnold et al. [2007] this effect was extended to show that disfluencies result in listeners' displaying a fixation bias toward an unfamiliar, hard to describe object over an every day object. The sensitivity which listeners display to the manner of an utterance's delivery can thus facilitate the processes involved in comprehension of the linguistic message by generating expectations about upcoming speech.

non-literal message Alongside influencing the interpretation of the linguistic message, listeners' biases to disfluencies also extend to their global pragmatic judgments made about an utterance. Estimates of the speaker's metacognitive state are sensitive to delivery, with utterances which are hesitant, disfluent, or with

rising intonation are judged as less confident [Smith and Clark, 1993; Brennan and Williams, 1995]. This effect has shown to be present in various contexts—in both speaker's own and a listener's assessment of certainty [Swerts and Krahmer, 2005], and for both adult and child listeners [Krahmer and Swerts, 2005].

The influence of paralinguistic features on listeners' judgments of a speaker's confidence is consistent with the effect of such cues on another pragmatic judgments—the recognition of deception. Listeners' accuracy in discriminating truth from lies has been shown to increase when presented with audio or audiovisual information rather than visual information [Bond and DePaulo, 2006], suggesting that vocal cues are more salient than visual in recognising dishonesty. Speaker-disfluency, and pauses in particular, appear to be frequently associated with an interpretation of dishonesty, in judgments made about speakers themselves as well as judgments made about other speakers [Zuckerman et al., 1981].

The effect of manner of delivery on listeners' pragmatic judgments have, until recently, used offline measures of judgement (post-hoc assessment of a speaker's confidence or honesty). However, the disfluency-dishonesty bias has also been shown to appear at the early stages of reference comprehension. Loy et al. [2016]looked at how listeners' pragmatic inferences about a speaker's honesty during online processing of speech was influenced by the presence of a disfluency. Subjects engaged in what was posed as a 'lie detection game', in which listeners were tasked to make implicit judgments about whether a speaker is being deceitful or not in indicating the location of a reward ("The treasure is behind the <referent>".). In a visual world paradigm, listeners signalled their assessment of a speaker's honesty by clicking on one of two objects (the referent and a distractor) - interpreting an utterance as honest would result in a click on the referent, and as dishonest a click on the distractor. Loy et al. [2016] found that subjects' eye and mouse movements displayed a bias toward the referred-to object for fluent utterances, and the distractor object for disfluent utterances.

As with the influence of disfluencies on listeners' predictions about semantic content [Arnold et al., 2004, 2007; Barr, 2001], manner of delivery's influence on listeners' pragmatic hypotheses emerges rapidly from the onset of critical words [Loy et al., 2016]. This finding is inconsistent with a traditional view [Blank, 1988] of comprehension which suggests that information about *how* an utterance is delivered (e.g. manner of delivery, social and contextual information), is only used to alter the global interpretation of an utterance *after* listeners have interpreted *what* was said.

associationist vs inferential/collateral The biases listeners' display to speaker-disfluency show evidence of an ability to to integrate information about utterance delivery during the moment-to-moment processing of speech. However, the mechanisms underlying how disfluencies cause listeners' expectations of upcoming speech to change are only recently becoming better understood. The low-level explanation is that listeners' sensitivity to utterance delivery results from a stochastic modelling of co-occurrences of disfluencies with certain properties of language use. For instance, listeners associating presence of disfluency with new information could be due to distributional learning of disfluencies more often occurring when speakers refer to something new than to something familiar [Arnold et al., 2004; Barr, 2001], and the disfluency-dishonesty bias in Loy et al. [2016] results from a prelearned association between disfluency and deception.

An alternative account suggests that paralinguistic features of speech are a form of collateral communication [Clark, 1996]. According to this view, listeners interpret disfluency as communicative signals aimed

at managing conversation. As opposed to having direct associations of disfluencies and language use, this view suggests that upon hearing a disfluency, listeners generate expectations by drawing inferences about its cause—i.e. about what might cause increased cognitive load for the speaker. This disfluency-attribution account of comprehension is paired with the suggestion that speakers' production of fillers is akin to the use of conventional words, intentionally signalling a delay to which the listener can then attribute a cause [Clark, 2002]. An account in which listeners engage in inferential processing about the plausible cause of speaker-disfluency implies that the online biases to disfluencies listeners have been shown to display will be speaker and situation specific. A listener's judgement on what might cause speaker-disfluency would be both speaker- and context-dependent.

heller, barr, arnold Recently, several studies have provided evidence for disfluency-comprehension being context-dependent by examining how listeners' biases to disfluencies change when presented with alternative plausible causes of speaker-disfluency. Arnold et al. [2007] showed that a bias towards unfamiliar, hard-to-describe objects following a disfluency is moderated by listeners' beliefs about that speaker. The results of both a gating task and an eye-tracking task showed that a) the discourse-new bias found in Arnold et al. [2004] could be extended to an unfamiliarity-bias, and b) this bias was reduced when the listener believed the speaker to have object-agnosia (inability to recognize objects). Similarly, Barr and Seyfeddinipur [2010] found that the discourse-new bias to disfluencies was dependent upon what was new for the speaker independent of what was new for the listener, and Heller et al. [2015] found that the bias towards unfamiliar objects extended to situations in which a

This shows that the effect which manner of delivery has on comprehension is dependent upon what is perceived to be the cause of the particular delivery, and is evidence of listeners' speaker-specific inferences affecting online comprehension - in this case the inference that both familiar and unfamiliar objects are equally difficult to describe for an object-agnosic speaker, affecting the listeners' online comprehension of disfluencies.

For listeners' perceptions about the cause of disfluency, belief that the speaker is object-agnosic provides an utterance-global explanation for speaker-disfluency. In this respect, whilst the integration of information evident in Arnold et al. [2007] clearly requires some form of inferencing, it is possible that listeners only engaged in the inferential process prior to speech onset. This could equate to essentially readjusting the statistical weight of the unfamiliarity-disfluency link, rather than repeatedly making inferential judgments about cause of disfluency as speech unfolds. The question remains whether listeners engage in inferential processing about the cause of disfluency *during* online speech processing.

what do we do? The current study develops Loy et al. [2016], and investigates whether the pragmatic deception-bias of disfluency is moderated by the presence of an utterance-local cause of speaker-disfluency - i.e. an explanation of disfluency which only presents itself *during* the time-course of an utterance. The idea is that if utterance-local contextual information informs the online pragmatic judgments made about disfluency, this would suggest that listeners' inferential processing in attributing a cause to a disfluency is a) flexible, and b) occurs during the moment-to-moment processing of speech. For the experiment reported here, we develop the 'lie detection task' from Loy et al. [2016] in which listeners were tasked to make implicit judgments about whether a speaker is being deceitful or not in indicating the location of a reward ("The treasure is behind the <referent>".). In a visual world paradigm, listeners signalled their assessment of a speaker's honesty by

clicking on one of two objects - interpreting an utterance as honest would result in a click on the referent, and as dishonest a click on the distractor. As well as manipulating the speaker's manner of delivery, utterances contained an alternative plausible cause of speaker-disfluency in the form of speaker-distraction. Utterances were presented to listeners under the guise of having been recorded outside in a busy street, and a strategically placed car-horn functioned as the speaker-distraction manipulation. Recording eye- and mouse- movements, we look at how the time course of pragmatic inferencing about speaker's disfluent message is influenced by the availability of an alternative, utterance-local cause of disfluency.

2 Method.

2.1 Materials.

Visual stimuli consisting of 120 black and white line drawings, taken from Snodgrass and Vanderwart [1980], were presented to participants in pairs across 60 trials (20 experimental, 40 fillers). Each trial presented the *referent* (the object which the speaker identified as having the treasure behind for that trial), and a *distractor*, which was chosen at random without replacement from a set of 60 objects. To control for the effect of the bias towards interpreting disfluency to difficulty of description [Arnold et al., 2007], critical referents and distractors were matched for familiarity (H value ≥ 3.0) and ease-of-naming (< 1.0). Object pairings with the same phonetic onset were avoided.

As in Loy et al. [2016] each referent was associated with a recording of a speaker naming the depicted object as the location of the treasure, and as before, critical utterances were either fluent or disfluent. The present study, however, manipulated not only the manner of delivery, but also the presence of a plausible cause of speaker distraction. Thus audio files were constructed such that the critical referents could be heard in four conditions in a 2 (fluent/disfluent) x 2 (distraction absent/present) design, see below. The insertion of a prolonged article followed by a filled pause into the fluent utterance formed the disfluent variants[1]. Meanwhile, to create a plausible cause of speaker-distraction, a 520ms car-horn sound effect was presented prior to the onset of the referent noun (-1100ms for disfluent utterances, -600ms for fluent).

1. **fluent, distraction absent:** The treasure is behind the <referent>.
2. **disfluent, distraction absent:** The treasure is behind thee, uh ... <referent>.
3. **fluent, distraction present:** The treasure is $\underbrace{\text{behind the}}_{\text{horn}}$ <referent>.
4. **disfluent, distraction present:** The treasure is $\underbrace{\text{behind thee}}_{\text{horn}}$, uh ...<referent>.

To situate the car-horn sound effect in a context, all recordings were presented overlaid upon a clip of ambient traffic and street noise which was unique to each referent. Thus the presentation of each critical referent varied only by 1) fluency and 2) presence of car-horn, to ensure that participants' behaviour was a reaction to the different combinations of these elements. All recordings, ambient noise and sound-effects were normalised and resampled to create 48 KHz stereo .wav files.

[1] We opted for utterance-medial disfluencies rather than utterance-initial as these sounded more believable as being caused by speaker distraction when paired with a environmental cue.

The 20 critical items/referents were counterbalanced across four lists, each with 10 fluent and 10 disfluent utterances, and each containing 10 instances of the car-horn (5 of which were disfluent). Lists also contained 40 filler utterances, half of which were fluent, and half of which contained either a form of disfluency or a discourse manipulation (see Table 3). Additionally, 20 of these filler items (10 fluent, 10 disfluent/discourse manipulation) contained various novel noises which could be interpreted as distracting for a speaker (see Table 4). These filler distractions varied in position relative to the referent noun onset.

2.2 Cover story.

Central to the idea of a distraction-caused-disfluency is the requirement that participants believe that the utterances were produced naturally and in a noisy environment. Participants were told that the recordings were made by a participant of a previous experiment which was conducted on the side of a busy street. To corroborate this, during the initial explanation of the experiment, participants were shown three videos claimed to be examples of the speaker producing the utterances. These were presented alongside the images about which the speaker spoke. Speaker utterances were depicted in the first video as honest & fluent, dishonest & disfluent in the second, and, in the third, honest & disfluent but distracted (speaker glances sideways toward a car following the car-horn sound effect).

To ensure that analysis could be run only on data from participants for whom the cover story held, a post-test questionnaire assessed whether participants thought it believable that the utterances were produced naturally and in the suggested context. This was double-checked by verbal questioning after participants were debriefed about the true nature of the experiment, by asking whether it occurred to them during the experiment that the recordings might not have been produced outside in the street.

2.3 Procedure.

Stimuli were displayed on a 21" CRT monitor, placed 850mm from an Eyelink 1000 Tower-mounted eye-tracker which tracked eye-movements at 500Hz (right-eye only). Audio was presented in stereo from speakers either side of the monitor. Mouse coordinates were sampled at 50Hz. The experiment was presented using OpenSesame version 3.0 [Mathôt et al., 2012].

Participants were told they would see two objects and hear a speaker identify one of them as hiding the treasure. They were told that the objective of the game was to accrue treasure by successfully locating the treasure, the location of which the speaker would lie half the time about. Participants' instructions were to click on the object which *they believed* the treasure to be behind. Once participants had read the instructions, the eye-tracker was calibrated. Calibration was monitored throughout, with the experiment being paused for recalibration if the experimenter deemed it necessary. To maintain accuracy, between each trial participants underwent a drift correction using a central fixation point, which changed from grey to red (for 500ms) upon successful fixation, signifying the start of the subsequent trial. Each trial began with the presentation of the two images (referent and distractor) in positions vertically central and horizontally to the left and right of centre (referents were presented equally often on each side). Simultaneous with presentation of the visual display was the onset of the ambient traffic audio file. After 1500ms, the mouse

pointer appeared centrally, and the playback of the utterance began. Upon clicking an object, the stimuli disappeared and were replaced by the grey fixation dot, signifying progression to the subsequent trial. Each trial had an automatic time-out of 5000ms from utterance onset, in case participants did not click on an object.

To maintain motivation throughout the study, participants were told that there were a number of "hidden bonus rounds" which offered more treasure. Of the filler trials, 25% did not immediately progress to the subsequent trial, instead giving a positive feedback message (regardless of which object was clicked) stating successful completion of a bonus round. Furthermore, participants were told that top scorers of the game would be able to enter their name on a highscore table, which was shown at the beginning of the experiment. Participants completed five practice trials (one of which was presented as a bonus round) prior to the main experiment. Eye-movements, mouse co-ordinates and object-clicked (referent/distractor) were recorded for each experimental trial.

3 Results.

3.1 Exclusion criteria.

For a final sample of 24, we were required to run the experient with 37 participants, recruited from the University of Edinburgh community, who participated in return for 4 remuneration. The data from 12 participants was excluded on the basis that they indicated either in the questionnaire (ten) or verbally (two) that they did not believe the cover story. Additionally, the study reported that one participant for whom the deception held had previously participated in Loy et al. [2016], and thus they were also exlcuded. Analysis was run on the data from the remaining 24 participants.

3.2 Analysis.

Analysis was carried out in R version 3.3.0 [R Core Team, 2016], using the lme4 package [Bates et al., 2015]. Trials in which participants did not click on either the referent or distractor (0.01%) were excluded from all analyses. Object-clicked (referent/distractor) was modelled using mixed effect logistic regression, testing for main effects of manner of delivery, presence of plausible speaker distraction, and their interaction as fixed effects, and including random slopes for these same variables, plus random intercepts and slopes by both subject and referent.

Eye-tracking fixation data was averaged into 20ms bins (of 10 samples) prior to analysis, and coded in terms of the area of interest (AOI), which were referent; distractor and neither image. The proportion of fixations to either object from the total sum of fixations was computed for each time bin. Mouse-tracking analysis only looked at the x-coordinates of the cursor. Distance traveled by the cursor (pixels) was computed for each 20ms bin, coded for direction relative to either object. The cumulative mouse distance traveled toward either object in each trial was calculated by summing the distance travelled up to each time bin, and proportion of movement was calculated by taking the distance travelled toward an object, divided by the total distance travelled in that trial in either direction. Trials for which the total mouse distance travelled post referent-onset was less than 1/3 of the distance from the screen centre to the near edge of an object

were excluded (0.03% of trials), on the basis that it was indicative of participants' having decided prior to referent-onset upon which object to click. Movements beyond the outer edge of either object were excluded on the basis that it indicated 'overshooting' with the cursor.

Model analysis for both eye- and mouse-movements were conducted over a time window from referent onset to 800ms post-onset. This window coincides with Loy et al. [2016], identified by evidence that reference is established in eye-movements around the 400-800ms post noun-onset mark [Eberhard et al., 1995], and covering the duration of the longest critical referent (776ms). Models were estimated using empirical logit regression [Barr, 2008] with main effects of time (scaled), object (referent/distractor), delivery (fluent/disfluent) and speaker distraction (absent/present), and interactions between them all, along with by-subject and by-item random intercepts and random slopes for time.

3.3 Final object-click.

Responses show the same overall tendency to interpret an utterance as truthful as was found in Loy et al. [2016], with 57% of trials resulting in a click on the referent and only 43% on the distractor. Participants were less likely to click on the referent following a disfluent utterance than a fluent one ($\beta = -2.24, SE = 0.67, p < .001.$). This is in keeping with literature [Brennan and Williams, 1995; Swerts and Krahmer, 2005] and with the results of Loy et al. [2016] - manner of delivery influences participants' global interpretation of the speaker's truthfulness. The presence of a plausible speaker distraction was not found to affect responses, neither was the interaction between delivery and distraction. The bias toward interpreting disfluency as a sign of dishonesty appeared to be explicit for 19/24 participants, as indicated in the post-test questionnaire. Table 2 shows the percentage of mouse-clicks on each object by condition.

3.4 Eye-movements.

The time-course (referent onset to 2000ms post-onset) of fixations to referent and distractor can be seen for the fluent conditions (Fig. 3) and the disfluent conditions (Fig. 4). Visualisation of the time-course of fixations show that for fluent utterances, participants displayed an early fixation bias towards the referent, which began to present itself prior to mean critical noun offset. For disfluent utterances, the picture is more complex—whilst a fixation bias towards the distractor was present, it appeared later on in the trial, not during the utterance of the critical noun as it did in Loy et al. [2016]. Although the same is true of utterances in which there was an available cause of speaker-disfluency, in these cases, participants appeared to fixate more to the referent first, before the later bias towards the distractor.

Model analysis of the window from referent-onset to 800ms post-onset confirmed what is evident from visualising the time-course of fixations. When presented with a fluent utterance, participants' fixations to the distractor decreased over time in favour of the referent ($\beta = -0.63, SE = 0.03, t = -18.51.$). In contrast, fixations to the distractor over the referent increased over time when the delivery was disfluent ($\beta = 0.59, SE = 0.05, t = 12.24.$). When a plausible cause of speaker distraction was present, participants' fixation biases towards the distractor during disfluent utterances were reduced ($\beta = -0.20, SE = 0.07, t = -2.9.$). The presence of the speaker-distraction was not found to have an effect on the tendency to fixate on the

referent for fluent utterances ($t = -0.9.$).

3.5 Mouse-movements.

The proportion of mouse movements (distance traveled) towards either object was computed for each 20ms bin from referent onset to 2000ms post-onset. Figures 1 and 2 and show the proportion of mouse movements towards either object in fluent and disfluent conditions respectively. Model analysis of the window from referent-onset to 800ms post-onset patterned with the eye-tracking data. When presented with a fluent utterance, participants' movements to the distractor decreased over time in favour of the referent ($\beta = -0.46, SE = 0.02, t = -19.49.$), whereas movements to the distractor over the referent increased over time when the delivery was disfluent ($\beta = 0.60, SE = 0.03, t = 17.72.$). When a plausible cause of speaker distraction was present, participants bias towards moving to the distractor during disfluent utterances was reduced ($\beta = -0.35, SE = 0.05, t = -7.25.$).

4 Discussion.

Evidence of listeners' pragmatic judgments about a speaker's honesty based on manner of delivery emerged at an early stage in utterance processing, with eye- and mouse-tracking data showing biases towards referent and distractor for fluent and disfluent utterances respectively during the initial stage of the comprehension process. In this respect, the present study successfully replicates the effect found by Loy et al. [2016], showing that an utterance's delivery influences listeners' pragmatic inferences about speaker's intentions to be (dis)honest. Additionally, the results here show that this effect is robust against the presentation of speech in a noisy environment, in which there are potential distractions for the listener.

The interest in this study, however, is in how listeners responded when presented with two competing explanations for the cause of speaker-disfluency. When an alternative explanation for speaker-disfluency was available to listeners, although the disfluency-dishonesty bias persisted in the final interpretation of an utterance (as evidenced by the eventual decision of which object to click), the picture during the initial stages of comprehension is more complex. The increase in fixations and mouse-movements to the distractor for disfluent conditions which occurred - indicating the emergent bias towards interpreting the utterance as dishonest - was initially attenuated when a possible cause of speaker-distraction was present. The availability of an external cause of the speaker disfluency resulted in a tendency to initially fixate and move toward the referent, only later fixating on and selecting the distractor.

One account of these findings would suggest that participants made early inferences about the contextual causes of disfluencies, which were eventually overridden by a dishonesty bias for disfluent utterances. By this account, listeners appear able to dynamically decide upon the most likely explanation of disfluency, and, as the speech unfolds, make attributions about why a speaker was disfluent. This equates to listeners engaging in a form of speaker/situation-modelling, in which judgments are made about a speaker's perspective and thought-processes. These judgments are then used to constrain the interpretation of the speaker's meaning$_{\text{nn}}$ and intentions. This evidence for perspective-taking in the early stages of reference comprehension goes against the view that adjustments for speaker, contextual and social constraints only emerge at later stages of

processing [Kronmüller and Barr, 2007].

However, it remains possible to explain these results in terms of direct associations of disfluencies and properties of language use. A probabilistic model of disfluency associations (e.g. deception, discourse-new or unfamiliar objects) would require a sensitivity to proximate environmental interference such that the weight given to other disfluency-biases is reduced. The distinction here is that disfluency effects on comprehension may still be a learned response to general language use, rather than the result of listeners' attributing beliefs and intentions to particular speakers. Such an explanation may, however, still require listeners to make some form of judgment about what contextual information justifies a reduction in a disfluency's influence—would the sensitivity listeners display to contextual information when processing disfluency discern between the car-horn used here, which was situated within the context of the speaker's environment, and the same car horn if it occurred in the street outside the lab and was audible to the listener but external to the experiment? It is hard to see how a judgment about what constitutes plausible speaker-distraction differs from an inference about the cause of a speaker's disfluency.

The study here shows that utterance-local contextual information influences listeners' moment-to-moment pragmatic inferences about disfluencies. Whilst being eventually overridden by listeners' bias towards interpreting disfluent utterances at dishonest, the availability of an alternative, external cause of speaker-disfluency (distraction), resulted in an initial attenuation of this bias. Whilst this contextual effect could be explained as a pre-learned expectation about disfluency frequently occurring after a disturbing noise, potential speaker-distractions are numerous and varied, and it is not clear how listeners might compensate for this. Alternatively, these findings would suggest that participants made early attributions about the potential causes of a disfluency, implying that listeners take into account a speakers perspective during the initial stages of comprehension.

References

Alibali, M. W., Heath, D. C., and Myers, H. J. (2001). Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen. *Journal of Memory and Language*, 44(2):169–188.

Arnold, J. E., Fagnano, M., and Tanenhaus, M. K. (2003). Disfluencies Signal Theee, Um, New Information. *Journal of Psycholinguistic Research*, 32(1):25–36.

Arnold, J. E., Kam, C. L. H., and 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., and Fagnano, M. (2004). The Old and Thee, uh, New. *Psychological Science*, 15(9):578–582.

Barr, D. and Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and cognitive processes*, 25(4):441–455.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication.

In Cavé, C., Guaïtella, I., and Santi, S., editors, *Oralité and gestualité: Interactions et comportements multimodaux dans la communication*, pages 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.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Beattie, G. W. (1979). Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation. *Linguistics*, 17(1-2):61–78.

Blank, G. D. (1988). Metaphors in the Lexicon. *Metaphors and Symbolic Activity*, 3(1):21–36.

Bond, C. F. and DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc*, 10(3):214–234.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, 44(2):123–147.

Brennan, S. and Williams, M. (1995). The Feeling of Anothers Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language*, 34(3):383–398.

Clark, H. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1):73–111.

Clark, H. H. (1996). Using language. 1996. *Cambridge University Press: Cambridge)*, 952:274–296.

Cook, S. W., Jaeger, T. F., and Tanenhaus, M. K. (2009). Producing Less Preferred Structures: More Gestures, Less Fluency. *31st Annual Conference of the Cognitive Science Society*, pages 62–67.

Corley, M. and Hartsuiker, R. J. (2011). Why Um Helps Auditory Word Recognition: The Temporal Delay Hypothesis. *PLoS ONE*, 6(5):e19792.

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

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., and Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24(6):409–436.

Fernald, A. and Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27(2):209–221.

Ferreira, F. and Bailey, K. G. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Sciences*, 8(5):231–237.

Fox Tree, J. E. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, 34(6):709–738.

Heller, D., Arnold, J. E., Klein, N., and Tanenhaus, M. K. (2015). Inferring Difficulty: Flexibility in the Real-time Processing of Disfluency. *Language and Speech*, 58(2):190–203.

Krahmer, E. and Swerts, M. (2005). How Children and Adults Produce and Perceive Uncertainty in Audiovisual Speech. *Language and Speech*, 48(1):29–53.

Kronmüller, E. and Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, 56(3):436–455.

Lau, E. F. and Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, 20(5):633–666.

Lickley, R. J. and Bard, E. G. (1998). When can listeners detect disfluency in spontaneous speech? *Language and speech*, 41 (Pt 2):203–226.

Loy, J. E., Rohde, H., and Corley, M. (2016). Effects of Disfluency in Online Interpretation of Deception. *Cognitive Science*.

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

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schachter, S., Christenfeld, N., Ravina, B., and Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, 60(3):362–367.

Smith, V. L. and Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, 32(1):25–38.

Snodgrass, J. G. and Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, 6(2):174–215.

Swerts, M. and Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53(1):81–94.

Zuckerman, M., Koestner, R., and Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, 6(2):105–114.

Did you find that these background noises were distracting for you as a listener?	58%
Did you think that these background noises affected the speaker?	50%
Did you think that the background noises, in some cases, might have caused the speaker to be disfluent (to say 'um' or 'uh')?	50%

Table 1: Percentage of participants who responded *yes/possibly/probably.*

Delivery	Disfluent	Disfluent	Fluent	Fluent
Speaker-distraction	Absent	Present	Absent	Present
Distractor	62%	62%	23%	24%
Referent	38%	38%	77%	76%

Table 2: Mouse clicks on each object by condition.

Filler type	Manipulation	No. of Utterances	Example
Fluent	None	20	The treasure is behind the <referent>.
Disfluent	Prolongation	3	The treasure is behind *thee...* <referent>.
	Repetition	4	The treasure is behind *the- the* <referent>.
	Filled pause (utterance-inital)	3	*Umm..* The treasure is behind the <referent>.
Other	Discourse marker	5	*Okay,* the treasure is behind the <referent>.
	Modal	3	The treasure *could be* behind the <referent>.
	Combination	2	*Right,* the treasure *might be* behind the <referent>.

Table 3: Disfluencies and discourse manipulations in filler items.

Noise	No. of Utterances
Vehicle horns (various)	7
Sirens (various)	3
Vehicles revving (various)	2
Car-stereo	2
Bicycle-bell	1
Bus doors opening	1
Footsteps	1
Loose drain cover	1
Man shouting	1
Dog barking	1

Table 4: Plausible causes of speaker distraction in filler items.

Effect	β	SE	t-value
(Intercept)	-0.60783	0.06074	-10.007
Time	0.33195	0.03640	9.118
Object	0.25363	0.03376	7.512
Delivery	0.09023	0.03384	2.666
Distraction	-0.11175	0.03377	-3.309
Time:Object	-0.62502	0.03376	-18.512
Time:Delivery	-0.21521	0.03384	-6.359
Object:Delivery	-0.32129	0.04785	-6.715
Time:Distraction	0.05362	0.03377	1.588
Object:Distraction	0.11787	0.04775	2.469
Delivery:Distraction	0.11588	0.04808	2.410
Time:Object:Delivery	0.58559	0.04785	12.238
Time:Object:Distraction	-0.04478	0.04775	-0.938
Time:Delivery:Distraction	-0.02308	0.04808	-0.480
Object:Delivery:Distraction	-0.02619	0.06797	-0.385
Time:Object:Delivery:Distraction	-0.19942	0.06797	-2.934
Random effects			
	Var.	S.D	Corr.
By-Subject			
(Intercept)	0.070	0.264	
Time	0.013	0.116	-0.720
By-Item			
(Intercept)	0.004	0.066	
Time	0.003	0.063	-0.940
Residuals			
	0.636	0.798	

Table 5: model eye window

Effect	β	SE	t-value
(Intercept)	-1.19815	0.03593	-33.35
Time	0.57152	0.03557	16.07
Object	0.17762	0.02370	.49
Delivery	0.22830	0.02400	9.51
Distraction	0.02378	0.02397	0.99
Time:Object	-0.46372	0.02379	-19.49
Time:Delivery	-0.28300	0.02407	-11.76
Object:Delivery	-0.31433	0.03392	-9.27
Time:Distraction	-0.08203	0.02400	-3.42
Object:Distraction	-0.04261	0.03388	-1.26
Delivery:Distraction	-0.04890	0.03411	-1.43
Time:Object:Delivery	0.60288	0.03402	17.72
Time:Object:Distraction	0.13227	0.03391	3.90
Time:Delivery:Distraction	0.21064	0.03411	6.18
Object:Delivery:Distraction	-0.09102	0.04821	1.89
Time:Object:Delivery:Distraction	-0.34941	0.04821	-7.25
Random effects			
	Var.	S.D	Corr.
By-Subject			
(Intercept)	0.020	0.141	
Time	0.016	0.128	0.110
By-Item			
(Intercept)	0.004	0.060	
Time	0.006	0.078	-0.790
Residuals			
	0.346	0.589	

Table 6: model mouse window

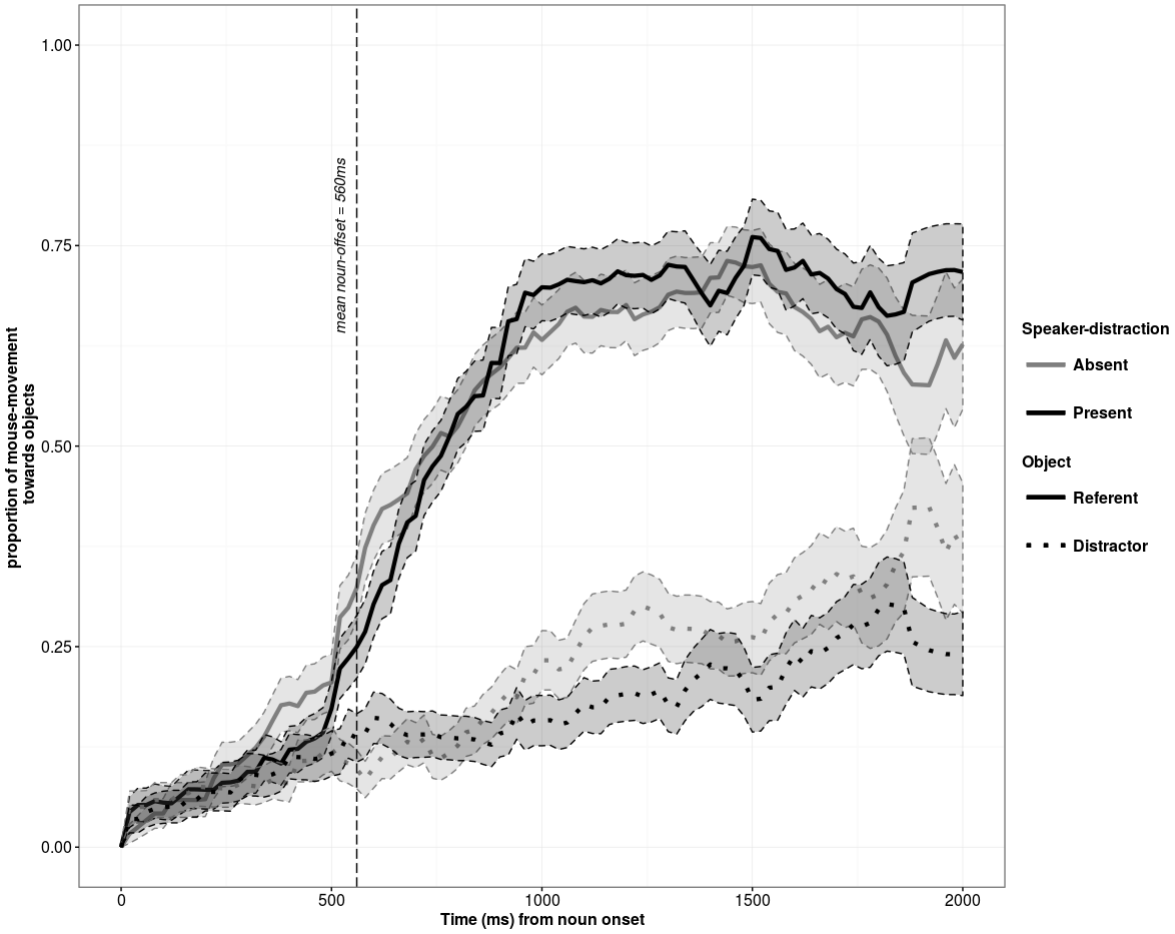

Figure 1: Mean proportion of cumulative distance traveled toward each object (referent or distractor) in fluent conditions split by presence of speaker distraction, from referent onset to 2000ms post-onset. Proportions calculated out of total cumulative distance moved the mouse from referent-onset until that time bin. Shaded areas represent ± 1 standard error of the mean.

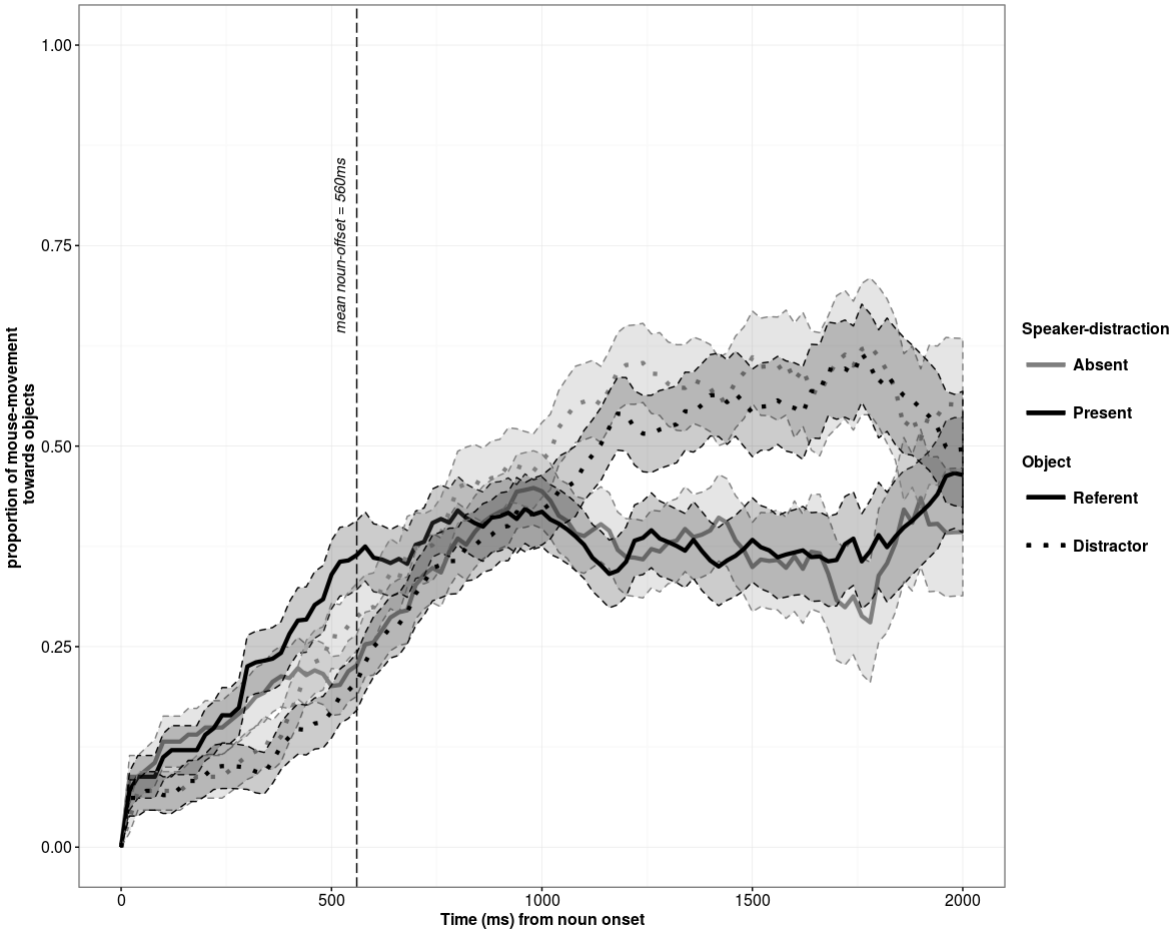

Figure 2: Mean proportion of cumulative distance traveled toward each object (referent or distractor) in disfluent conditions split by presence of speaker distraction, from referent onset to 2000ms post-onset. Proportions calculated out of total cumulative distance moved the mouse from referent-onset until that time bin. Shaded areas represent ±1 standard error of the mean.

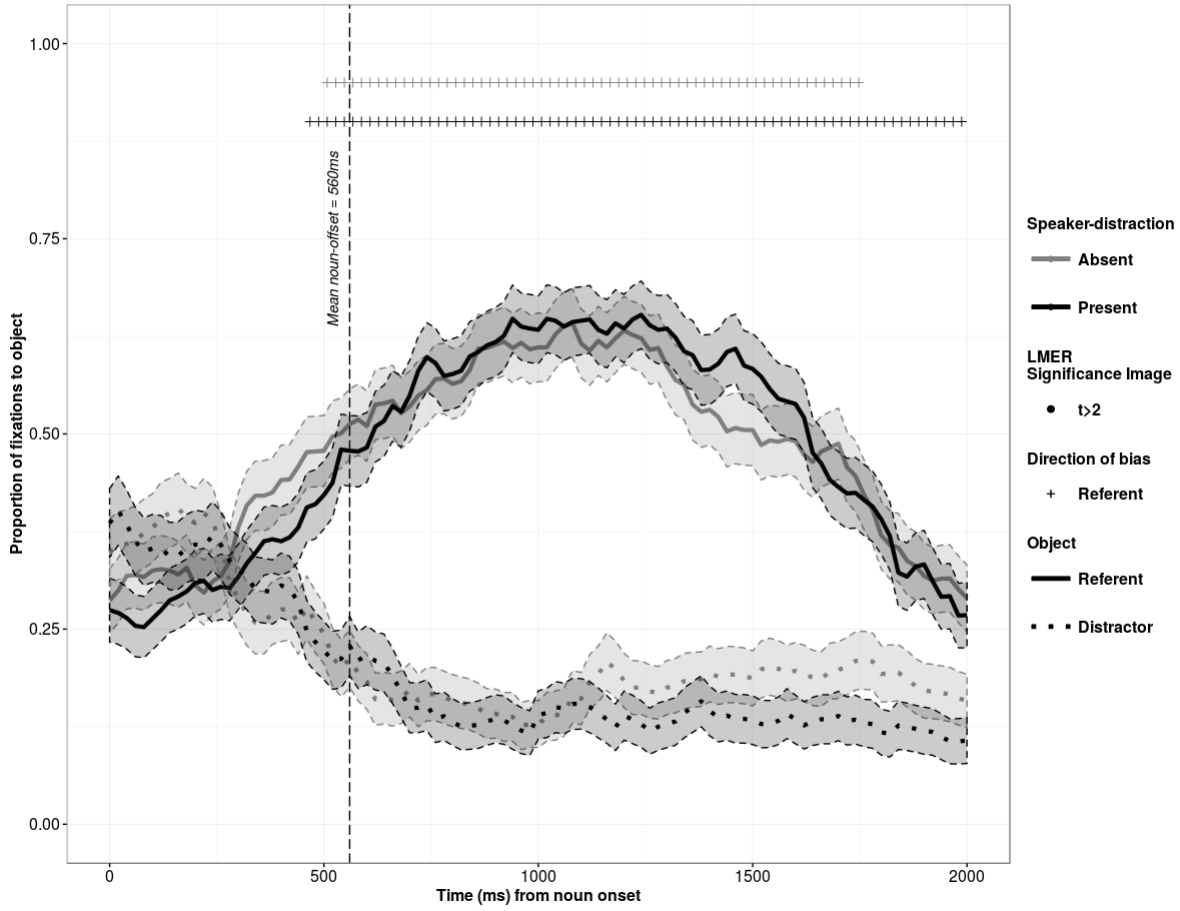

Figure 3: Mean proportion of fixations to either object (referent and distractor) for fluent utterances split by presence of speaker distraction, calculated out of the total sum of fixations for each 20ms time bin from referent-onset to 2000ms post-onset. Shaded areas represent ±1 standard error of the mean. Time bins in which a bias towards either the referent or distractor was evident are highlighted with respect to condition and direction of bias (assessed by a mixed effect model estimating the empirical logit of fixations to either object in a given condition, with main effect of object and by-subject and by-item random intercepts and random slopes of object.

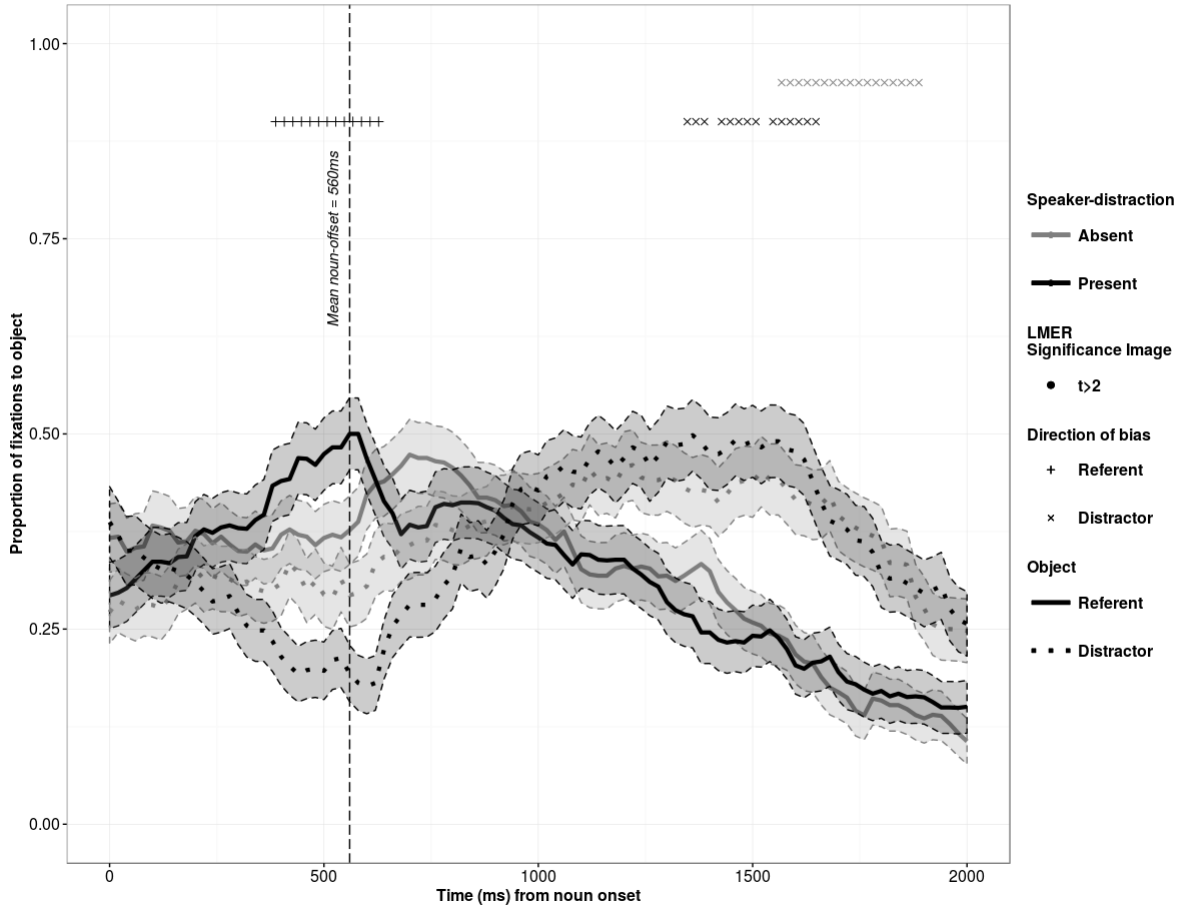

Figure 4: Mean proportion of fixations to either object (referent and distractor) for disfluent utterances split by presence of speaker distraction, calculated out of the total sum of fixations for each 20ms time bin from referent-onset to 2000ms post-onset. Shaded areas represent ±1 standard error of the mean. Time bins in which a bias towards either the referent or distractor was evident are highlighted with respect to condition and direction of bias (assessed by a mixed effect model estimating the empirical logit of fixations to either object in a given condition, with main effect of object and by-subject and by-item random intercepts and random slopes of object.