Effects of Disfluency in Online Interpretation of Deception

Jia E. Loy

Psychology, PPLS, University of Edinburgh

Hannah Rohde

Linguistics and English Language, PPLS, University of Edinburgh

Martin Corley

Psychology, PPLS, University of Edinburgh

Author Note

Martin Corley

PPLS, 7 George Square,

Edinburgh EH8 9JZ, UK

Martin.Corley@ed.ac.uk

Abstract

A speaker's manner of delivery of an utterance can affect a listener's pragmatic interpretation of the message. Disfluencies (such as filled pauses) influence a listener's off-line assessment of whether the speaker is truthful or deceptive. Do listeners also form this assessment during the moment-by-moment processing of the linguistic message? Here we present two experiments that examined listeners' judgment of whether a speaker was indicating the true location of the prize in a game during fluent and disfluent utterances. Participants' eye and mouse movements were biased toward the location named by the speaker during fluent utterances, whereas the opposite bias was observed during disfluent utterances. This difference emerged rapidly after the onset of the critical noun. Participants were similarly sensitive to disfluencies at the start of the utterance (Experiment 1) and in the middle (Experiment 2). Our findings support recent research showing that listeners integrate pragmatic information alongside semantic content during the earliest moments of language processing. Unlike prior work which has focused on pragmatic effects in the interpretation of the literal message, here we highlight disfluency's role in guiding a listener to an alternative non-literal message.

Keywords: Psychology; Language understanding; Pragmatics

Effects of Disfluency in Online Interpretation of Deception

Introduction

In interpreting an utterance, listeners must often take into account not only *what* is said, but *how* it is being said, because the manner of production may in turn be relevant to the interpretation of meaning. For instance, a speaker may place a prosodic contour on a statement to emphasize a point (Fernald & Mazzie, 1991), or they may preface their utterance with an "um" to express a lack of certainty (Smith & Clark, 1993). Such paralinguistic features encoded within the speech signal are known to influence listeners' pragmatic interpretations of the linguistic message. In their study on the perception of speaker confidence, Brennan and Williams (1995) observed that utterances produced with rising intonation, temporal delays, and filled pauses such as "uh" or "um" were judged to be less confident, showing that the manner of delivery of an utterance provides cues that influence a listener's estimation of the speaker's metacognitive state. Brennan and Williams' (1995) findings have since been extended to a variety of contexts, such as the assessment of a speaker's certainty (Swerts & Krahmer, 2005), by both adult and child judges (Krahmer & Swerts, 2005), as well as in machine interpretation of nonspontaneous speech (Pon-Barry & Shieber, 2011).

Research on the recognition of deception also highlights the perceptual relevance of paralinguistic information in language comprehension. Studies on lie detection focus on the interpretation of paralinguistic cues in both auditory and visual modalities. Vocal cues, such as tone of voice or speech disturbances, appear to be more informative than visual cues such as facial gestures or body movements. Across an analysis of 50 studies that manipulated participants' access to visual and audio information when making truth-lie judgments, Bond and DePaulo (2006) found that discrimination accuracy was lower when participants made judgments based on visual information rather than audio or audiovisual, highlighting the salience of vocal cues in the context of recognizing deception.

Filled pauses, in particular, stand among a set of belief cues that listeners frequently associate with lie perception. Zuckerman, Koestner, and Driver (1981) found that speech hesitations were one of the cues reliably associated with judgment of lie-telling, whether it was in predicting behavior exhibited by subjects themselves or by others. This finding corroborates a parallel assessment of the actual cues that listeners use in detecting deception, where a meta-analysis of 33 studies again implicated filled pauses as an indicator of deception (Zuckerman, DePaulo, & Rosenthal, 1981).

The aforementioned studies thus outline a relationship between a speaker's manner of delivery and the listener's pragmatic interpretation of the speech; however, one limitation in the existing research is the reliance on off-line measures of judgment. While the time-course of pragmatic inferencing lies beyond the scope of those studies, off-line methods are consistent with an implicit assumption that in computing the meaning of an utterance, contributions made by pragmatic enrichment occur at a relatively late stage of processing. This is in line with traditional models of language comprehension, which frequently distinguish between the what and the how of a linguistic message. Under this view, listeners must first interpret the literal message content before consulting any social or contextual knowledge that may alter the global meaning of the utterance (Blank, 1988; Hamblin & Gibbs, 2003; Keysar, Barr, Balin, & Brauner, 2000; Lattner & Friederici, 2003). Pragmatic considerations, such as what might be intended by the utterance or whether the speaker might be lying, thus require additional (and correspondingly time-consuming) inferential processes before a contextually-appropriate meaning can be derived (e.g., Blank, 1988).

This temporal prioritization of the literal, however, has not gone unchallenged. A growing body of research emphasizes the primacy of the communicative aspect of linguistic exchange (Clark, 1996). Studies on figurative language processing suggest that people are not necessarily slower in interpreting the figurative meaning of an expression than the literal (e.g., Blasko & Connine, 1993; Gibbs, 1994; Giora, 1997; see Glucksberg, 2003 for a review). Using a cross-modal priming paradigm, Blasko and Connine (1993) measured the reaction times for participants making lexical decisions on words that were literally-related, metaphorically-related or unrelated to the vehicle of a metaphor that they heard, and showed that people access metaphorical meanings as quickly as the literal, even in the case of novel or unfamiliar metaphors such as “her thoughts were a boiling kettle”. Given adequate contextual information, listeners can comprehend the intended meaning of non-literal utterances directly without first having to compute and reject a contextually-incompatible literal interpretation (Gibbs, 1994; cf. Pynte, Besson, Robichon, & Poli, 1996).

Findings such as these are inconsistent with traditional models, highlighting the problematic nature of the “literal-first” assumption. On the other hand, their conclusions converge with several studies examining the time course of pragmatic inferencing which have found immediate effects of various kinds of contextual constraints on listener interpretation (e.g., Nadig & Sedivy, 2002; Hanna, Tanenhaus, & Trueswell, 2003; Hanna & Tanenhaus, 2004; Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; although cf. Lattner & Friederici, 2003). For instance, results from event-related brain potentials (ERPs) reveal that listeners make inferences based on the speaker’s voice, and what it suggests about age and gender. Evidence for these inferences emerges as early as 200-300 ms from the onset of a critical word, suggesting that pragmatic inferences are processed by the same early interpretation mechanisms that construct a content-based literal meaning (Van Berkum et al., 2008). Van Berkum et al. interpret their findings within a one-step model of language comprehension, where social context is incorporated immediately and alongside semantic content in the construction of an overall utterance meaning.

Eye-tracking studies on perspective-taking show that addressees are able to take into account common ground information available to both themselves and the speaker from the earliest moments of reference resolution (Hanna et al., 2003; Nadig & Sedivy, 2002). Hanna and Tanenhaus (2004) showed that listeners interpreting an ambiguous referent were sensitive to the contextual constraint of whether or not the speaker was physically able to reach the object, with effects emerging rapidly after the onset of the critical object name. Studies such as these show that listeners are able to rapidly extract and use pragmatic information based on the speaker’s identity, and based on the situation that the utterance is produced in; two forms of context to which the listener has access prior to and during the entire utterance. Building on this, the experiments reported here are designed to establish whether listeners make rapid pragmatic inferences based on the moment-to-moment manner in which an utterance is spoken. Specifically, we investigate how a listener's on-line judgment of whether an utterance is a truth or lie varies with the voice-based cue of utterance fluency, exploiting the well-established finding which indicates a correlation between disfluencies and lie perception.

To date, the majority of research on the on-line comprehension of disfluencies has focused on the effect of disfluency on expectations relating to the semantic content of the message (Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Arnold, Hudson Kam, & Tanenhaus, 2007; see Arnold & Tanenhaus, 2011 for a review). For instance, Arnold et al. (2004) showed that eye movements of listeners told to manipulate objects on a visual display were biased toward either previously-mentioned or discourse-new objects, depending on whether the instruction contained a disfluency. Listeners hearing "Put the grapes above the candle. Now put the candle/camel…" were initially more likely to look at the candle, whereas disfluent instructions such as "Now put thee, uh, candle/camel…" elicited more looks toward the camel during the temporarily ambiguous onset ("ca…"). Arnold et al. (2007) later extended this finding to unfamiliar objects perceived as difficult to describe, by demonstrating that listeners hearing instructions such as "Click on thee, uh, green…" were initially more likely to look toward green squiggly objects than green everyday items such as ice cream cones. These studies provide evidence that listeners are sensitive to the manner of delivery when predicting the upcoming semantic content of the message; disfluencies create probabilistic expectations about which object a speaker is likely to refer to out of a given set of possible objects.

In a subsequent experiment, Arnold et al. showed that this semantic effect of disfluency is modulated by factors such as prior knowledge about the speaker. When they informed listeners that the speaker giving the instructions had object agnosia, they found that the tendency to fixate difficult-to-name objects was sharply reduced, highlighting the context-dependent nature of the disfluency bias (Arnold et al., 2007, Experiment 2). Barr and Seyfeddinipur (2010) observed a similar effect of disfluency, where listeners' content-based predictions were modulated by their inferences relating to the speaker's familiarity with the object. Using a mouse-tracking paradigm, they showed that disfluencies facilitated mouse movements toward a target when described by a speaker who had not seen the object before, but not during descriptions by a speaker familiar with the object. These studies show that various forms of contextual information can influence a listener's interpretation of disfluency when making on-line predictions about the message content. However, it remains unclear from the current literature how disfluencies may affect a listener's pragmatic hypotheses during the moment-to-moment processing of the linguistic message itself.

In the current study we aimed to investigate whether, and how, the presence of disfluency in an utterance influences listeners' pragmatic inferences relating to the speaker's truthfulness during on-line comprehension of the linguistic message (i.e., an inference about an alternative non-literal message). In the experiments reported here, we presented a speaker as being sometimes dishonest within a context framed as a lie detection task. Using a game in which listeners were led to assess whether or not a speaker was indicating the true location of a reward, we manipulated the speaker's manner of delivery (fluent vs. disfluent) to test how this influenced whether or not the listener believed the utterance. We made use of an eye- and mouse-tracking paradigm to establish the time course with which these pragmatic inferences are made. Together, the two measurements provided a picture of the focus of listeners' visual attention through their eye movements (Altmann & Kamide, 2007) and the continuous trajectory of decision-making through their mouse movements (Spivey & Dale, 2006). In Experiment 1, we used an utterance-initial filled pause to create the disfluent stimuli, in line with previous studies on manner of delivery (e.g., Brennan & Williams, 1995); in Experiment 2, we explored whether the location of disfluency matters by shifting the disfluency to an utterance-medial position. As we will show, the results demonstrate that listeners are influenced by both utterance-initial and utterance-medial disfluencies when making an implicit judgment on the veracity of the speaker's utterance. Eye- and mouse-tracking data also suggest that this inference is made during early moments of utterance comprehension—during fluent utterances, listeners were quickly biased toward the object referenced by the speaker, whereas with disfluent utterances they were biased toward the other object.

Experiment 1

Experiment 1 was designed to test whether a speaker's manner of delivery influences a listener's interpretation of an utterance as a truth or lie, as well as whether this judgment is formed during on-line language comprehension. We told participants that they were taking part in a lie detection study. The experiment was presented in the format of a computer game, with a series of opportunities for the participant to uncover hidden treasure behind one of two visible objects on the screen. Participants were told that the goal was to accumulate treasure by clicking on the object that concealed the treasure on each trial. Participants heard a recorded speaker reference which object to click on, but were informed beforehand that the speaker was a participant from an earlier experiment who had been instructed to lie half the time about the treasure's location. This served to establish an element of potential deception in the experiment, as well as to justify the presence of disfluency in the stimuli. We analyzed the participants' eye and mouse movements as well as their object clicks.

Method

Participants. 21 self-reported native speakers of English took part in the experiment. Participants were all right-handed mouse users with normal or corrected-to-normal vision.

Materials. Visual stimuli comprised 120 line drawings from Snodgrass and Vanderwart (1980), presented in pairs across sixty trials (20 critical; 40 filler). For each pair presented in a visual display, we will use the term *referent* for the object that the speaker named as the object concealing the treasure; we will refer to the other object as the *distractor*. Critical referents and distractors were matched for ease of naming (H

value < 1.0) and familiarity (≥ 3.0) to minimize participants' biases toward either object based on expectations relating to difficulty of description (cf. Arnold et al., 2007).[1] Care was also taken to ensure both objects did not start with the same sound on critical trials (cf. Arnold et al., 2004).

Each referent was associated with a recording specifying the image as the object that the treasure was hidden behind. Critical utterances were either fluent (*The treasure is behind the ...*) or were preceded by an utterance-initial filled pause (*Um, the treasure is behind the ...*). The sentences were initially recorded in their entirety; a 400 ms filled pause from one of the disfluent utterances was then cross-spliced onto the start of each fluent statement to create a corresponding disfluent version. This ensured that participants were reacting to the same utterance (bar disfluency manipulation) in each condition, and to the same disfluency across all disfluent trials.

The 20 experimental referents were counterbalanced across two lists each containing ten fluent and ten disfluent utterances, such that each referent that occurred within a fluent utterance in the first list occurred within a disfluent utterance in the second. Each list included an additional 40 filler utterances, also naming a referent as the object concealing the treasure. To increase variability, half of these included one of various forms of disfluency, or a discourse manipulation such as a non-propositional sentence marker or a modal varying the speaker's commitment to the truth value of the statement. The reason for the filler manipulation was twofold: (a) to distract participants from the filled pause manipulation on critical trials, and (b) to create a set of utterances that closely approximated natural speech in order to reinforce the cover story which emphasised that sentences were unscripted. A summary of filler utterance types is provided in Table 1. The remaining 60 objects served as distractors. These were randomly paired with referents on each display, with no repetition of images across the experiment.

[1] H is a measure of name agreement for a given concept or image, where higher values correspond to greater difficulty experienced by speakers in naming the object. A description of its derivation can be found in Snodgrass and Vanderwart (1980), from which the H values and familiarity ratings reported here were taken.

Procedure. The experiment was presented using OpenSesame version 2.9.4 (Mathôt, Schreij, & Theeuwes, 2012) on a 21in. CRT monitor. Eye movements were monitored using an Eyelink 1000 Tower Mount system which tracked the right eye, sampling at 500Hz. Mouse coordinates were sampled at 50 Hz.

Fig. 1 presents a sample of a trial from the experiment. Between trials, participants saw a central gray fixation dot and underwent a manual drift correct to ensure accurate recordings from the eye tracker. After this, the fixation dot turned red for 500 ms to signify the start of the trial. The dot disappeared and was replaced by two images (a referent and a distractor), which were centered vertically and positioned horizontally left and right on the screen. Referents appeared on the left and right side equally often. Each pair of images appeared for a 1000 ms preview, after which the mouse pointer appeared at the center of the screen and playback of the audio stimulus began. Participants were instructed to click on the object which they believed concealed the treasure. Once a mouse click had been recorded on one of the two objects, the objects disappeared and the gray fixation dot appeared to begin the next trial, except in the case of fillers which included feedback (see below). Trials had an automatic time-out 5 seconds post-audio onset. If a click was not detected before this, participants saw a message telling them to respond more quickly.

To keep participants motivated, we informed them that the game contained a number of hidden "bonus" rounds which offered more treasure than the average trial. To simulate these rounds, 25% of filler trials were programmed to display a message that a treasure chest had been found regardless of which object was clicked on. This message appeared immediately following the detection of a mouse click on an object, and remained on screen until participants clicked again to begin the next trial. Additionally, participants were told that top scorers would be able to enter their name into the high score table, which appeared on the screen at the start of the experiment. A 5-trial practice session preceded the main experiment. One of the practice trials was always set to display the bonus treasure message. After the task, a post-experiment questionnaire was used to verify that none of the participants suspected that the audio stimuli comprised speech that had been scripted for purposes of the experiment.

For each experimental trial, we recorded eye and mouse movements as well as which object (referent or distractor) the participant clicked on.

Results

Trials on which a click was not recorded on either object were excluded from analysis (0.5% of experimental trials). Statistical analyses were carried out in R version 3.1.0 (R Development Core Team, 2015) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). We modelled the binary outcome of final object click in a mixed effect logistic regression model to test for the main effect of manner of delivery (fluent/disfluent), including random intercepts and slopes for subject and item random effects. Eye-tracking records were averaged into 20 ms bins, each comprising ten samples, prior to analysis. Fixation data were coded in terms of region of interest (referent/distractor/none), and the proportion of fixations to each object 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 traveled by the mouse was computed by taking the difference between the X coordinates of the current and previous sample. The data were coded for direction of movement (toward referent/toward distractor) for each bin, and the cumulative distance participants had moved the mouse toward either object was calculated by summing over the distance traveled in each direction up until that time bin (taking into account all previous mouse movements in that direction on that trial). For each object, we then calculated a proportion-of-movement measure, defined as the distance traveled by the mouse pointer towards the given object, divided by the total distance traveled (regardless of X direction).

In all figures, the proportion of fixations or mouse movements to each object is plotted from onset of the referent (the point of disambiguation in the utterance) until 2000 ms post-onset, by which point participants had typically moved the mouse over one of the two objects. Model analyses for eye and mouse movements were conducted over a time window beginning from referent onset to 800 ms post-onset. This window was identified based on existing research which suggests listeners' eye movements establish reference around 400 to 800 ms after an object is mentioned (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Hanna et al., 2003; Tanenhaus, 2007). The window extends over the mean duration of critical referents (560 ms) and ends just after offset of the longest critical referent (776 ms), to minimise possibility of confounding results with any post-processing effects. Models were estimated using empirical logit regression (Barr, 2008), including time and manner of delivery (fluent/disfluent) as fixed effects, and random intercepts and random slopes for time in by-subjects and by-items analyses.

Final object click

The distribution of responses suggests an overall tendency to believe the speaker to be truthful rather than deceptive. Across the experiment, clicks on the referent were recorded in 58% of trials while clicks on the distractor were recorded in 42% of trials. Table 2 shows the breakdown of mouse clicks on each object in each condition. Participants were more likely to click on the referent following a fluent utterance than a disfluent one, $\beta = 2.30$, SE = 0.48, $p < .001$. Manner of delivery therefore appeared to influence perception of the speaker's truthfulness such that participants were more inclined to interpret a fluent utterance as truthful and a disfluent one as deceptive.

Eye movements

We calculated for each 20 ms time interval the proportion of fixations to each object, out of the total sum of fixations (including those to areas that constituted neither object). Fig. 2 shows the proportion of fixations to each object in each condition until 2000 ms post-noun onset. The time course of fixations reveals an effect that emerges rapidly in the fluent condition, with looks to the referent rising and diverging from the distractor beginning about 300 ms post-noun onset. In the disfluent condition, we initially see a rise in looks toward the referent, before an abrupt decrease around 600 ms post-onset and a corresponding increase in looks to the distractor. The initial rise in fixations to the referent likely reflects eye-movement sensitivity to establishing reference when an object is named (e.g., Eberhard et al., 1995); this explains the slight delay with which effects emerge during disfluent utterances relative to fluent.

To test how the effect of manner of delivery changed over time, we calculated the empirical logit of the fixation proportion to the referent for each 20 ms time bin over our 0 to 800 ms window of interest, averaging across all trials in each condition separately for subjects and for items. Model analysis revealed a reliable interaction between time and manner of delivery, $\beta = 1.72$, SE $= 0.7$, $t = 2.47$ by subjects; $\beta = 1.01$, SE $= 0.39$, $t = 2.58$ by items, reflecting the difference in listener commitment toward the object mentioned by the speaker dependent on manner of delivery.

Mouse movements

Mouse movements were scored according to whether they were in the direction of the referent or the distractor, and the absolute distance traveled (in pixels) was computed for each 20 ms time step. Fig. 3 shows the proportion of mouse movements (in terms of distance traveled) toward each object until 2000 ms post-noun onset. Examination of the total distance traveled by the mouse from the time of noun-onset to the point of response suggests an object bias dependent on manner of delivery. On fluent utterances, participants' cursors were more strongly attracted to the referent than the distractor, while disfluent utterances exhibited the opposite trend. This effect was confirmed by a model analysis on total mouse distance which revealed an interaction between object and manner of delivery, $\beta = 253.48$, SE $= 87.67$, $t2.89$ by-subjects; $\beta = 211.80$, SE $= 89.15$, $t = 2.38$ by-items.

The time course of mouse movements demonstrates that this difference emerged quickly post-noun onset. On fluent utterances, mouse movements toward the referent begin rising sharply and diverging from movements toward the distractor 400 ms post-onset. On disfluent utterances, movements reflect a small but distinct preference toward the distractor from around the same time, although a prominent divergence only appears about 1500 ms post-onset (see Fig. 3). Comparing movements toward the referent in both conditions though, it is clear that participants were more likely to move the mouse toward the referent during fluent utterances. This effect emerged 500 ms after onset of the noun, confirming that the disfluency affected listeners' interpretation during early moments of comprehension. Following the eye-tracking analysis, we computed the empirical logit of the proportion of distance traveled by the cursor toward the referent for each 20 ms time interval, calculated out of the total distance traveled in either direction. Results indicated a time by manner of delivery interaction for both subjects, $\beta = 7.47$, SE $= 2.91$, $t = 2.56$, and items, $\beta = 3.47$, SE $= 1.50$, $t = 2.30$. As with eye movements, participants' mouse movements appear rapidly sensitive to the manner of delivery such that fluent utterances are more likely to elicit movements toward objects named by the speaker compared to disfluent utterances.

Discussion

Experiment 1 established that listeners make pragmatic inferences about a speaker's truthfulness dependent on the manner of delivery of the utterance. Both eye- and mouse-tracking results indicate that this bias emerged over a time window corresponding to early moments of utterance processing. Although the time course suggests a slight temporal delay in mouse movements relative to the eye, this is unsurprising as it merely reflects the role of the visual system in the planning and execution of hand motor motions (cf. Land, Mennie, & Rusted, 1999). Accordingly, listeners' eye movements exhibited an initial sensitivity to the object being named during disfluent utterances, but a corresponding effect of mouse movements toward the referent was not observed. More importantly, the pattern of bias in the time-course of eye and mouse movements follow a similar trend. Fluent utterances yielded an increase in fixations and mouse movements to referent objects (i.e., participants were inclined to interpret the statement as truthful), while the opposite trend was observed for disfluent utterances. Taken together, these results indicate a clear effect of manner of delivery on the listener's interpretation of the speaker's truthfulness. Importantly, they suggest that there is rapid integration of pragmatic information during on-line comprehension, allowing participants to quickly assign a non-literal meaning to the speaker's mention of a referent. The speed with which they do this suggests that the disfluency has already been incorporated into their model of the speaker and has an immediate effect on their interpretation of the utterance.

Experiment 2

Experiment 2 was designed to test whether listeners' pragmatic inferences about speaker veracity are similarly sensitive to utterance-medial disfluencies. Previous production studies on disfluency suggest that utterance-initial disfluencies are linked to global, speech-planning issues. For example, Clark and Fox Tree (2002) observed that ums occurred most frequently at the start of intonation units, which they attributed to greater demands on the speech-planning mechanisms (cf. Swerts, 1998; Watanabe, 2002). Utterance-medial filled pauses, on the other hand, tend to be associated with localized, lexical access difficulties (e.g., Beattie & Butterworth, 1979). The design of comprehension studies to date have largely aligned with these accounts. Studies examining listeners' global, metacognitive perceptions of speech such as those relating to a speaker's state have mainly made use of utterance-initial filled pauses (e.g., Brennan & Williams, 1995), while those investigating listeners' expectations relating to the semantic content of the message have focused on the comprehension of utterance-medial pauses (e.g., Arnold et al., 2004). Hence, we tested whether the presence of disfluency mid-utterance affects listeners' global judgment of whether the speaker was lying or telling the truth. Based on findings from Experiment 1, if utterance-medial disfluencies similarly influence listeners' perception of deception, we would expect to see temporal evidence of this effect reflected in participants' early eye and mouse movements.

Method

Participants. 22 participants took part in Experiment 2, all of whom fulfilled the same requirements as Experiment 1. None had participated in Experiment 1.

Materials. The same objects were used in Experiment 2 as in Experiment 1. Disfluent stimuli for both experiments were recorded during the same session to ensure consistent prosody with filler utterances. As with Experiment 1, the disfluent segment was cross-spliced from one of the disfluent utterances into each fluent utterance to create a disfluent counterpart. Disfluent utterances were characterized by an utterance-medial disfluent segment comprising a prolonged article followed by a filled pause (*The treasure is behind thee, uh . . .*). The 20 critical referents were counterbalanced across two lists each containing ten fluent and ten disfluent utterances. Filler utterances characterized by an utterance-medial filled pause in Experiment 1 were replaced accordingly for an utterance-initial filled pause. The rest of the filler utterances remained the same as Experiment 1.

Procedure. The experiment was identical in every respect to Experiment 1, with the only difference being the disfluency position in the recordings. A post-experiment questionnaire was again used to check whether any participants were sensitive to the experimental manipulations. Data from one subject who guessed that the speech was scripted was excluded from the analysis.

Results

We followed the same analysis procedures in Experiment 2 as in Experiment 1. Data from 4 experimental trials (0.9%) on which a click was not recorded on either object were excluded from the final dataset.

Final object click

The response distribution again suggests a global bias toward the object named by the speaker: Participants clicked on the referent in 57% of trials and on the distractor in 43% of trials. Table 2 shows the distribution of trials on which participants clicked on either object in each condition. As in Experiment 1, this was influenced by the speaker's manner of delivery, with participants more likely to click on the referent following a fluent utterance, $\beta = 4.06$, SE $= 0.60$, $p < .001$. This replication of results from Experiment 1 suggests that listeners are sensitive to utterance-medial disfluencies when making pragmatic inferences about a speaker's truthfulness. To test whether there were differences in sensitivity to utterance-initial and utterance-medial disfluencies, we compared final object clicks for the subset of disfluent trials in both experiments. Participants were less likely to click on the referent following utterance-medial disfluencies, $\beta = -1.52$, SE $= 0.52$, $p < .005$. We return to this effect in the General Discussion.

Eye movements

Fig. 4 shows the proportion of fixations to each object in each condition until 2000 ms post-noun onset. The effect of manner of delivery is reflected in eye movements as the speech unfolded over time. During fluent utterances, fixations to the referent rose quickly, accompanied by a corresponding decrease in looks to the distractor. Conversely, disfluent utterances yielded the reverse pattern. The time course of events is comparable to that observed in Experiment 1, with a fixation bias emerging 300 ms after onset of the referent in the fluent condition, and around 600 ms post-onset in the disfluent condition. This effect on looks to the referent over time was reflected in a significant interaction between time and manner of delivery, for both subjects and items analyses on the time window $0 - 800$ ms post-noun onset, $\beta = 3.82$, SE $= 1.33$, $t = 2.86$, and $\beta = 2.96$, SE $= 0.95$, $t = 5.01$. A comparison of eye movements during disfluent utterances across the two experiments found no effect of disfluency location, $\beta = -0.26$, SE $= 0.32$, $t = -0.81$.

Mouse movements

Mouse-tracking results reveal a contrast between fluent and disfluent utterances which corroborates the eye-tracking. Analysis of the distance traveled toward each object as a proportion of the total distance participants moved the mouse suggests a bias due to manner of delivery: on fluent utterances, participants' cursors were overwhelmingly more likely to be attracted to the referent, while disfluent utterances were characterized by more movements toward the distractor. This difference was confirmed by a significant interaction between object and manner of delivery, $\beta = 475.95$, SE = 96.95, $t = 4.91$ by subjects; $\beta = 377.35$, SE = 70.33, $t = 5.37$ by items.

Fig. 5 shows the proportion of mouse movements (in terms of distance traveled) toward each object until 2000 ms post-noun onset. Time course examination demonstrates that the object bias emerged rapidly after onset of the disambiguating noun, with a referent-bias beginning 300 ms post-onset during fluent utterances, and a distractor-bias beginning 200 ms during disfluent ones, showing that participants' mouse movements were quickly sensitive to the speaker's manner of delivery in their decision of which object to click on. As in Experiment 1, model analysis taking the empirical logit of the proportion of distance traveled toward the referent as the dependent variable yielded a significant interaction between time and manner of delivery, both by-subjects, $\beta = 11.04$, SE = 2.69, $t = 4.10$, and by-items, $\beta = 6.73$, SE = 2.82, $t = 2.39$. As for eye movements, there was no effect of disfluency location on mouse movements when comparing Experiments 1 and 2, $\beta = 2.22$, SE = 2.90, $t = 0.76$.

Post-hoc analyses

We conducted two subsidiary analyses to rule out potential counterexplanations of our findings. The first examined whether participants who knew they might be deceived might become sensitive to the distributions of fluent/disfluent utterances, using them as discriminative cues due to nature of the task. The second considered whether disfluencies were understood as a general signal of uncertainty, rather than as a specific signal of deception.[2] To maximise power, we present analyses from data pooled across Experiments 1 and 2, although we note that the findings from analyses of individual experiments do not differ.

In order to establish whether participants learned to discriminate the disfluent signal, we compared the trials encountered during the first third of each experiment to

[2]We thank a reviewer of an earlier version of this paper for these suggestions.

those encountered during the last third. There was no effect of trial subset for either eye movements, $\beta = -0.09$, SE $= 0.30$, $t = -0.29$, or mouse movements, $\beta = -0.17$, SE $= 0.48$, $t = -0.36$, confirming that the results were unlikely to be a learned response due to participants picking up on experimental contingencies over time.

To test whether disfluencies were percieved as general signals of uncertainty (cf. Brennan & Williams, 1995), we analyzed only those items containing a disfluency, comparing trials that resulted in a judgment of truthfulness versus deception. Figs. 6 and 7 show that participants' eye and mouse movements demonstrate an early object bias contingent on the final response judgment (based on the object that they ultimately clicked on for that trial); trials which resulted in a judgment of truthfulness saw a referent bias, while those resulting in a judgment of deception saw a distractor bias. This difference was reflected in a time by object interaction for both fixations, $\beta = 2.23$, SE $= 0.29$, $t = 7.61$, and mouse movements, $\beta = 12.32$, SE $= 3.02$, $t = 4.08$, confirming that the effect of disfluency on listeners' inferences about the speaker's truthfulness occurred during the initial stages of comprehension.

General Discussion

The distribution of responses in both experiments indicates an overall tendency to believe the speaker was being truthful rather than deceptive. While this result differs from what we might expect by chance, it follows a trend observed in previous studies on proposition verification, which provide evidence to suggest an *a priori* prejudice in listeners toward believing something to be true (e.g., Barres & Johnson-Laird, 2003; McKinstry, Dale, & Spivey, 2008). These studies show that when asked to assess the truth value of a statement, participants demonstrate greater difficulty in evaluating a statement as false, suggesting an expectation regarding at least a partial adherence to the Gricean maxim of quality on the speaker's part. In the current study, we observed that this pre-existing "truth-bias" still appears to bear weight even when listeners have been told to expect a potentially non-Gricean, uncooperative speaker. Notably, this trend also aligns with findings from actual lie detection studies, which point at a general tendency toward perceiving speakers to be truthful rather than deceptive. For example, Vrij (2000) found a mean of 61.5% truth judgments across nine studies analyzing listeners' truth-lie discrimination accuracy. This lends face value to the suggestion that the behaviour of participants in the current study was comparable to that of judges discriminating between authentic truths and lies.

Extending previous findings on the effect of manner of delivery, our results show that the way in which a message is conveyed affects a listener's implicit judgment of the speaker's truthfulness. Specifically, fluent utterances bias listeners toward perceiving a speaker to be truthful, as evidenced by increased fixations on, and mouse movements towards, the object named by the speaker, while disfluency biases listeners toward expecting the speaker to lie. This is consistent with previous studies on deception, which show that listeners believe liars to be more disfluent than truth-tellers. Together, the two experiments establish that listeners' pragmatic interpretations are affected very quickly by manner of delivery, whether disfluency occurs utterance-initially or mid-utterance.

Utterance-medial disfluencies resulted in fewer clicks on the named referent than did utterance-initial disfluencies. With the caveat that this is a between-experiment comparison, this difference appears difficult to reconcile with the production-based view that lies are planned at the level of the message, and that any difficulty associated with lying should therefore be best evidenced by utterance-initial disfluencies. This highlights the complexities associated with extending production accounts of disfluency to listener comprehension. While speakers may be disfluent in ways that are predictable, listeners' sensitivities to disfluencies might depend on multiple factors, such as their ability to model the mind of the speaker, or the contexts in which these cues arise. For instance, a phrase-medial filled pause may arise as a consequence of a speaker's search for a word; however, in the context of anticipating a potentially dishonest speaker, the listener may interpret the disfluency as a pragmatic cue to deception. Such a disjunct between speakers' productions and listeners' perceptions of disfluencies has been recently shown in the context of a competitive game involving deception (Loy, Rohde, & Corley, 2016).

Contrary to predictions made by traditional models of language comprehension, temporal information from the eye- and mouse-tracking provides little evidence to suggest that pragmatic inferences are relegated to a later, post-literal stage of processing. Rather, our results support recent research showing that pragmatic information is extracted and used as an immediate constraint on utterance processing (e.g., Hanna & Tanenhaus, 2004; Nadig & Sedivy, 2002). Recall that in order to isolate effects during on-line comprehension, our time course analyses were conducted over a window beginning from noun-onset and ending 800 ms post-onset. Given that the average duration of critical nouns in the experiment was 560 ms (range: 413 – 776), when taking into account the 200 ms it typically takes to program an eye-movement (Matin, Shao, & Boff, 1993), it is clear that participants' eye and mouse movements began to exhibit this bias as the critical linguistic input unfolded. Thus, pragmatic judgments about the veracity of the speaker's utterance were made on-line, and modulated immediately by the presence or absence of a paralinguistic cue.

Building on previous studies which show that disfluency affects listeners' on-line expectations about the literal message content, our findings demonstrate that disfluency modulates listeners' pragmatic hypotheses about the speaker's intentions simultaneously alongside integration of lexical, semantic information. The speed with which these hypotheses unfold is consistent with the disfluency bias observed by previous studies, delineating an immediacy associated with judgment biases triggered by paralinguistic information such as filled pauses. Within the framework of detecting deception, this immediacy is also compatible with longstanding stereotypes of deceit that people seem to retain, despite repeated findings that perceivers tend to perform at or near chance level at overt lie detection tasks (e.g., ten Brinke, Stimson, & Carney, 2014). Given the ingrained nature of these stereotypes, it is perhaps less surprising that such biases are hard to overcome, whether in the context of judging deception or some other pragmatic aspect of the utterance in question.

One question arising from the results here is why the distractor-bias observed in the disfluent condition is smaller than the referent-bias in the fluent condition. One possible explanation could be due to the nature of the experimental stimuli, which resulted in fluency being made salient in the presence of disfluent utterances. Consequently, participants may have made their judgment based on the speaker's relative fluency (rather than simply treating disfluency as a systematic indicator of deception), resulting in what could be explained as an effect of fluency in response to disfluency. However, because participants had a pre-existing bias toward expecting the speaker to be truthful, this created a baseline preference for the object identified by the speaker. This in turn had to be overcome in order for participants to choose the distractor object during utterances perceived as lies, resulting in a greater referent-bias during fluent utterances compared to the corresponding distractor-bias during disfluent utterances. This raises the question of whether there are probabilistic cues to deception (such as disfluency in speech) that listeners implicitly rely on to recognize a deceptive speaker, or whether the process is in fact modulated by contextual reasoning (such as perception of an unexpected change in the speaker's manner of delivery). If the latter were the case, a disfluent utterance may be less likely to be perceived as a lie when produced by a hesitant speaker compared to one that was consistently fluent. Exploring how listeners integrate various sources of information when making pragmatic inferences would be a useful avenue for future research.

To our knowledge, this is the first study to integrate eye- and mouse-tracking measures in a visual world paradigm. It would appear that they corroborate each other, with the eyes tending to fixate the target that listeners are moving the mouse towards. In addition to validating conclusions from previous studies that mouse coordinates can serve as a temporally sensitive index of language processing comparable with eye movements (Farmer, Anderson, & Spivey, 2007; Spivey, Grosjean, & Knoblich, 2005), this opens up opportunities for further studies seeking to employ a mouse-tracking only methodology (e.g. in the case of certain clinical or developmental populations, which can sometimes pose challenges when eye-tracking; cf. Sasson & Elison, 2012), or even studies combining the two to more fully understand the perceptual-motor processes underlying spoken language comprehension.

The current results show that listeners make use of information in a disfluency to draw pragmatic inferences about a speaker's truthfulness, and more importantly, that this information is brought to bear during initial stages of the comprehension process. These results are relevant not only to comprehension in contexts which involve deliberate deception, but also in the many contexts in which a listener believes a speaker to be cooperative but must still take care to consider the status of the literal message—is it true? Is additional meaning intended? (e.g. "your haircut looks, um, great"; "the house for sale is, um, cozy"). In this way, we propose that the immediacy of listeners' pragmatic use of disfluency in evaluating an utterance is applicable broadly in language comprehension.

References

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Arnold, J. E., Hudson Kam, C. L., & 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. (2011). Disfluency effects in comprehension: How new information can become accessible. In E. Gibson & N. J. Pearlmutter (Eds.), *The processing and acquisition of reference* (pp. 197–217). Cambridge, MA: MIT Press.

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.

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.

Barres, P., & Johnson-Laird, P. N. (2003). On imagining what is true (and what is false). *Thinking & Reasoning*, *9*(1), 1–42.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=lme4` (R package version 1.1-6)

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*(3), 201–211.

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

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 295–308.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

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.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111.

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

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, *57*(4), 570–595.

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

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge: Cambridge University Press.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96.

Hamblin, J. L., & Gibbs, R. W. (2003). Processing the meanings of what speakers say and implicate. *Discourse Processes*, *35*(1), 59–80.

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.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(1), 29–53.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing - evidence from event-related brain potentials. *Neuroscience Letters*, *339*(3), 191–194.

Loy, J., Rohde, H., & Corley, M. (2016, May). Lying, in a manner of speaking. In *Proceedings of Speech Prosody 8.*

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.

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22–24.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.

Pon-Barry, H., & Shieber, S. M. (2011). Recognizing uncertainty in speech. *Eurasip Journal on Advances in Signal Processing.*

Pynte, J., Besson, M., Robichon, F. H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, *55*(3), 293–316.

R Development Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

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

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 25–38.

Snodgrass, J. G., & 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 and Memory*, *6*(2), 174–215.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211.

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.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*(4), 485–496.

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

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. , 309–26.

ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some evidence for unconscious lie detection. *Psychological Science*, *25*(5), 1098–1105.

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

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and implications for professional practice*. New York: Wiley.

Watanabe, M. (2002). Fillers as indicators of discourse segment boundaries in Japanese monologues. In *Proceedings of speech prosody* (pp. 691–694).

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in experimental social psychology*, *14*, 1–59.

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

Table 1

Breakdown of filler stimuli and examples of each type.

Filler type	Manipulation	No. of utterances	Example
Fluent	None	20	The treasure is behind the giraffe.
Disfluent	Prolongation	3	The treasure is behind *thee...* mushroom.
	Repetition	4	The treasure is behind *the- the* roller skate.
	Filled pause (utterance-medial)	3	The treasure is behind the, *uh* rooster.
Other	Discourse marker	5	*Okay,* the treasure is behind the clothes peg.
	Modal	3	The treasure *could be* behind the balloon.
	Combination	2	*Right,* the treasure *might be* behind the caterpillar.

Table 2

Breakdown of mouse clicks recorded on each object (referent or distractor) by manner of delivery (fluent or disfluent) for Experiments 1 and 2. Values represent percentage of trials.

		Experiment 1	Experiment 2
Fluent	Referent	78	87
	Distractor	22	13
Disfluent	Referent	37	17
	Distractor	63	83

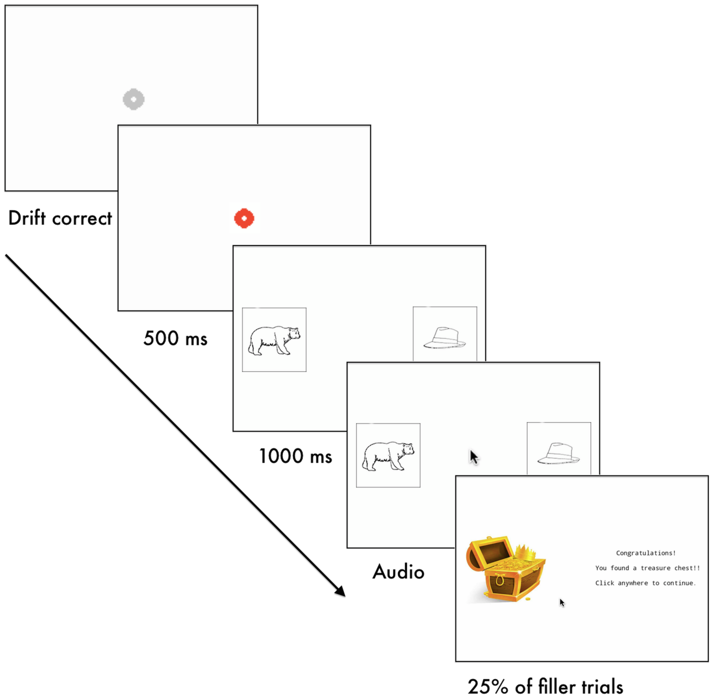

Figure 1. Timeline of a sample trial from the experiment

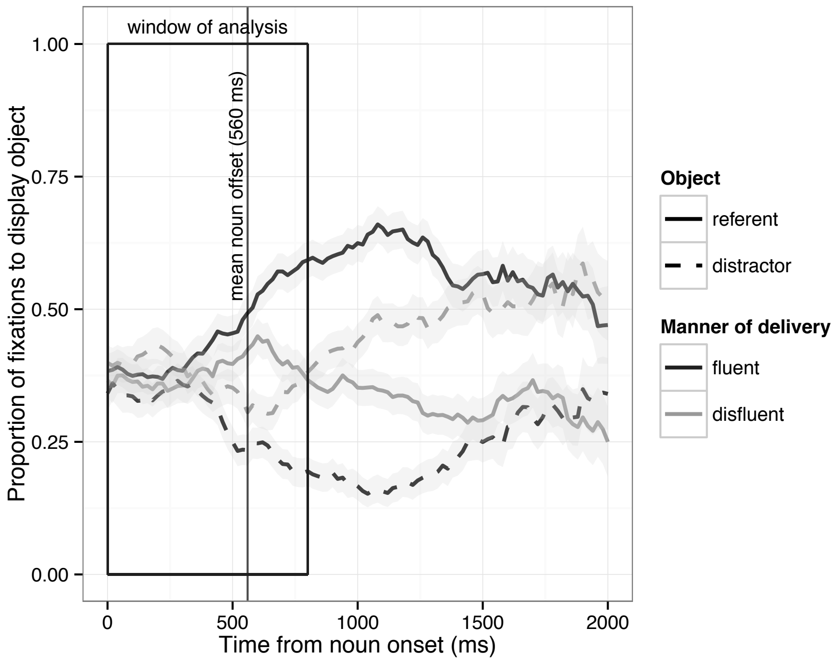

Figure 2. Eye-tracking results for Experiment 1: Proportion of fixations to each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ±1 standard error of the mean. Note that proportions do not sum to 1 because some fixations fell outside of either object.

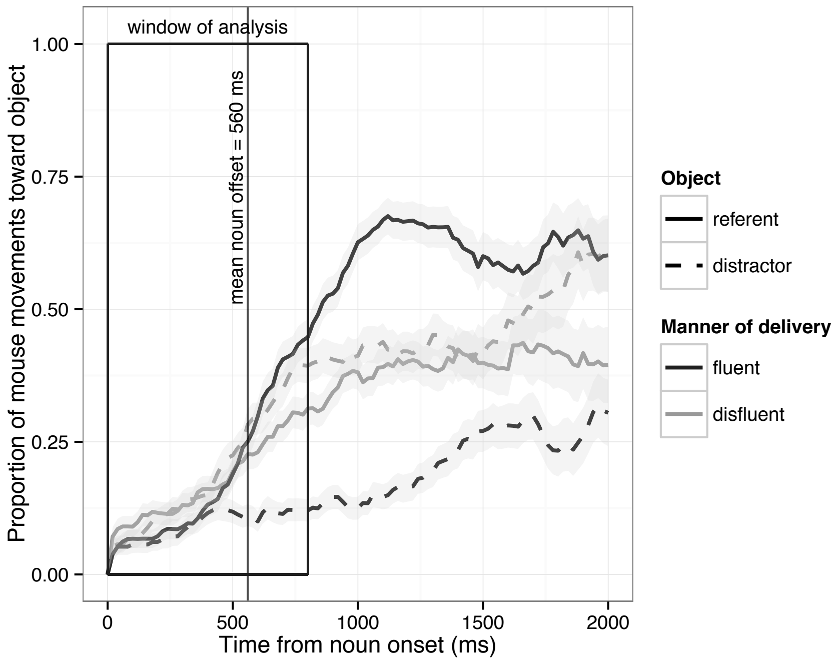

Figure 3. Mouse-tracking results for Experiment 1: Proportion of cumulative distance traveled toward each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset. Proportions were calculated out of the total cumulative distance participants moved the mouse from noun-onset until that time bin. Shaded areas represent ±1 standard error of the mean.

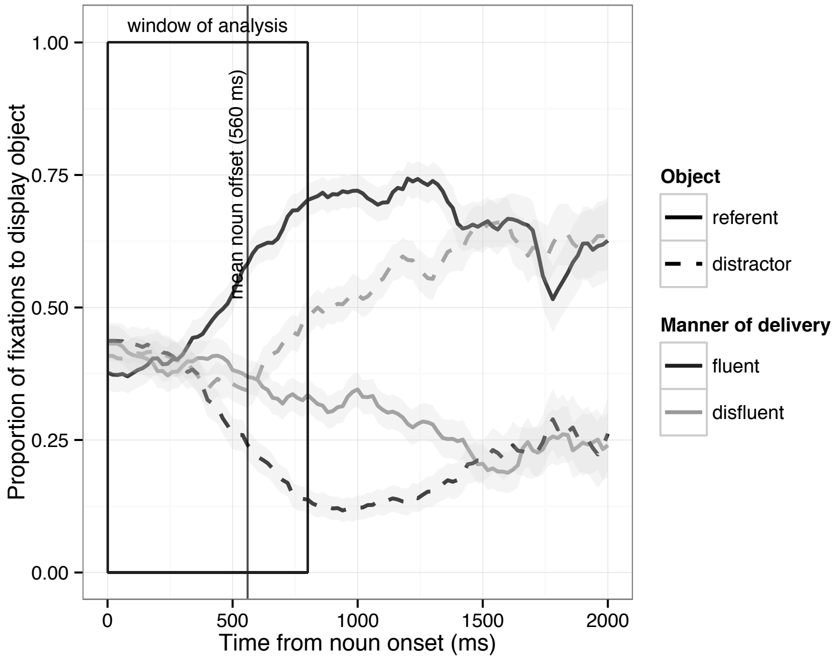

Figure 4. Eye-tracking results for Experiment 2: Proportion of fixations to each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ±1 standard error of the mean. Note that proportions do not sum to 1 because some fixations fell outside of either object.

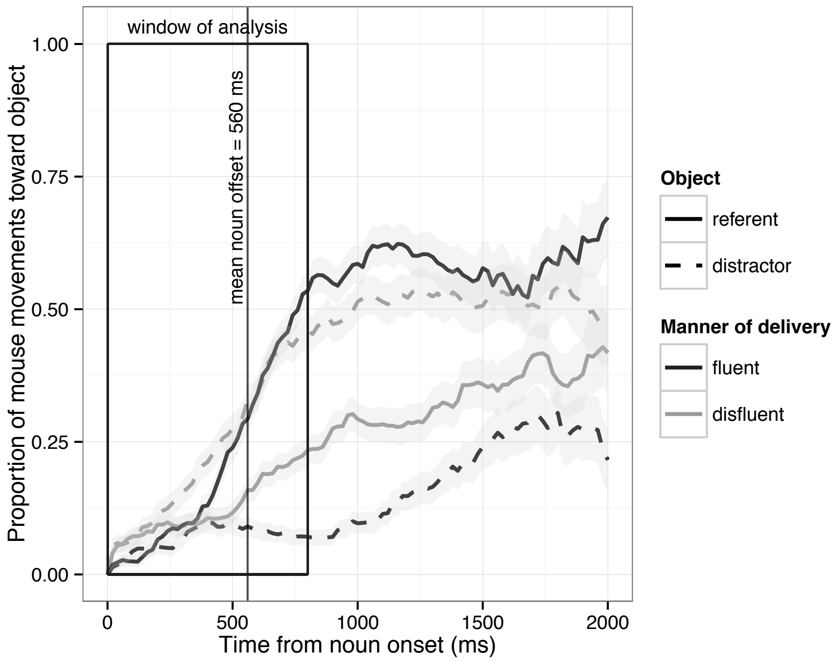

Figure 5. Mouse-tracking results for Experiment 2: Proportion of cumulative distance traveled toward each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset. Proportions were calculated out of the total cumulative distance participants moved the mouse from noun-onset until that time bin. Shaded areas represent ±1 standard error of the mean.

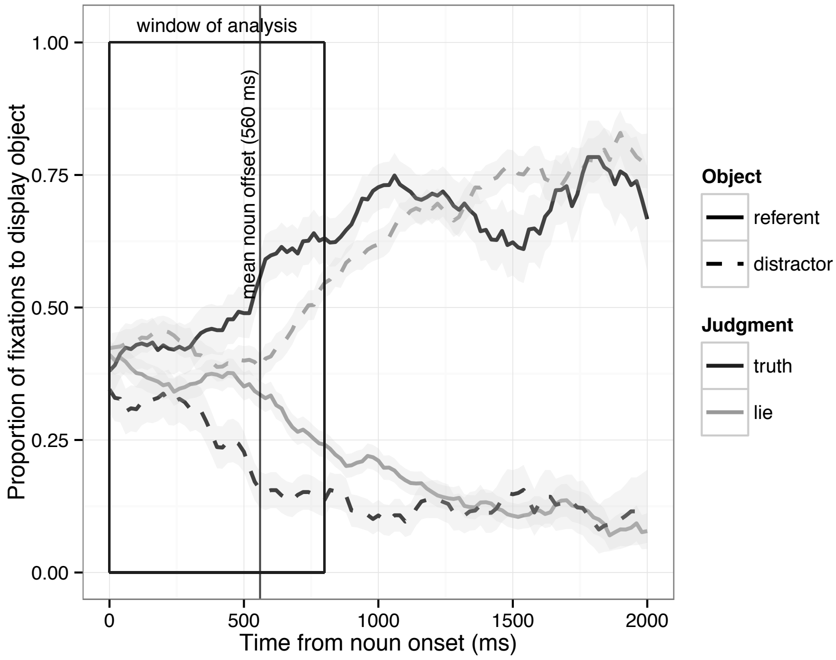

Figure 6. Eye-tracking results for subset of disfluent trials from Experiments 1 and 2: Proportion of fixations to each object (referent or distractor) broken down by response judgment (truth or lie), from 0 to 2000 ms post-noun onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ±1 standard error of the mean.

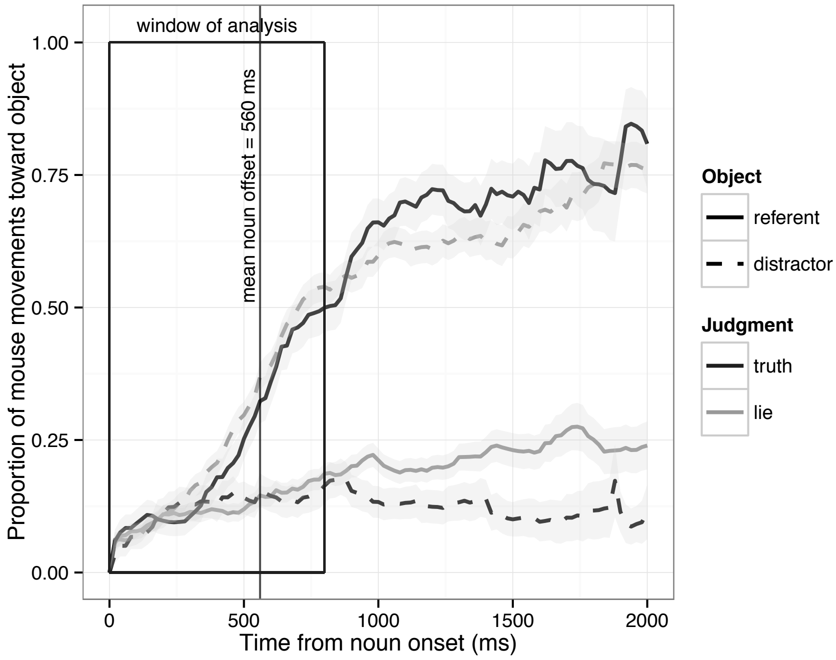

Figure 7. Mouse-tracking results for subset of disfluent trials from Experiments 1 and 2: Proportion of cumulative distance traveled toward each object (referent or distractor) broken down by response judgment (truth or lie), from 0 to 2000 ms post-noun onset. Proportions were calculated out of the total cumulative distance participants moved the mouse from noun-onset until that time bin. Shaded areas represent ±1 standard error of the mean.