Such gesture, very lie, wow

reorder(MC,JK,JL)

Psychology, PPLS, University of Edinburgh

Josiah King

Philosophy, Psychology and Language Sciences

University of Edinburgh

7 George Square

Edinburgh EH8 9JZ, UK

J.P.J.King@sms.ed.ac.uk

Abstract

Previous research suggests that, when questioning the veracity of an utterance, we perceive certain non-verbal behaviours to indicate that a speaker is being deceptive. Recently, work has highlighted how listeners' associations between speech disfluency and dishonesty happen at the earliest stages of reference comprehension, suggesting that contextual information about the manner of spoken delivery influences pragmatic judgments simultaneously with integration of lexical information. The studies presented here ask whether this listeners can also rapidly integrate the visual channel in making judgments about a speaker's honesty. Eye- and mouse-tracking experiments investigate the time-course of the associations listeners' might have between different types of gestures and deceit. Participants saw and heard a video of a potentially dishonest speaker describe treasure being hidden behind a named object, while also viewing both the named object and a distractor object. Their task was to click on the object behind which they believed the treasure to actually be hidden. Experiment 1 investigates which, if any, types of gestures listeners might associate with deception. Experiment 2 establishes a dishonesty-bias for adaptor gestures, and shows that this begins at an early stage in reference comprehension. Additionally, we examine whether listeners' judgments of deception correlate with how nervous they think the speaker appears in each video. These experiments highlight the potential obstacles to incorporating gestures into the visual world paradigm, and ultimately show the value of doing so. Results show that the integration of the visual modality can have a rapid and direct influence on pragmatic judgments, supporting the need to study communication as a multimodal phenomenon.

Such gesture, very lie, wow

That people deceive one another is an inescapable aspect of everyday communication. After studying participants' social interactions over a one week period, DePaulo, Kashy, Kirkendol, Wyer, and Epstein (1996) concluded that people lie on average twice a day. The idea that it possible to detect deceit — that there are systematic differences in people's behaviour depending on whether they are telling a truth or a lie — has long captivated human interest in a variety of areas, from criminal interrogations to the business world.[1] Research suggests, however, that the behaviours which we associate with lying are often at odds with those behaviours which actually signal dishonesty; a better understanding of why and how listeners interpret certain non-verbal behaviours as 'cues to deception' can help to explain this inconsistency. The present experiments investigate listeners' associations between a speaker's honesty and their non-verbal behaviour, and study the time-course of how multi-modal information (speech and gesture) is integrated influence such pragmatic judgments.

Much research has investigated the behaviours which listeners associate with a speaker's honesty. In speech, vocal cues such as disfluencies are frequently shown to signal dishonesty (Vrij, Kneller, & Mann, 2000; Zuckerman, DePaulo, & Rosenthal, 1981). These associations occurring during the early stages of comprehension, almost as soon as listeners can infer whether deception has occurred (Loy, Rohde, & Corley, 2017). Interestingly, listeners appear to hold this association despite inconclusive evidence to indicate its accuracy as a strategy for detecing deceit: A meta-analysis of 116 studies by DePaulo et al. (2003) found little evidence for speech disturbances as cues to deceit, and some production studies even suggest that the opposite association of lies being more fluent than truths may be more accurate (e.g. Arciuli, Villar, & Mallard, 2009; Benus, Enos, Hirschberg, & Shriberg, 2006). The frequency and speed with which listeners continue to associate disfluency with deception highlights its robustness as a cue to perceived deceit.

[1] Where, in some cases, "how to lie" is equally as important. See https://www.marketplace.org/2008/02/18/busness/lying-essential-doing-business

In most natural communication, however, speakers can convey information via multiple channels, all of which may influence pragmatic judgements of their (dis)honesty: Lies may be discernable from truths in body-language and gestures as well as in spoken delivery. The influence of producing a lie on a speaker's body movements may pattern with its influence on their speech: A reduction in illustrative gesturing has been associated with deceit (Cohen, Beattie, & Shovelton, 2010; DePaulo et al., 2003), as has both shorter talking time and a reduction in detail (DePaulo et al., 2003), suggesting that speakers are less informative — in both speech and gesture — when lying. Just as speech contains disfluencies, the visual modality contains a collateral channel in the form of potentially unconscious movements and postures. The validity of visual cues as *actual* signals of deception is inconclusive. The two main meta-analyses in the area have mixed conclusions: Zuckerman, DePaulo, and Rosenthal (1981) found more shrugs and fidgetting be indicative of deceit, whereas DePaulo et al. (2003) found little evidence of a relationship between lying and most forms of movement. Some studies even suggest that deception is associated with a decrease in hand, arm, and leg movements (e.g. DePaulo, 1992; Ekman, 1989; Vrij, 1995).

Just as there appears to be a disparity between listeners' beliefs about what a liar sounds like and how liars actually speak, beliefs about what a liar looks like appear to be at odds with how liars actually behave. Across 33 studies, Zuckerman, DePaulo, and Rosenthal (1981), found that nine out of ten visual cues were *believed* to be associated with deceit, but only three reliably signalled actual deception. Furthermore, in a subset of 13 studies which reported relationships between potential cues and subsequent deception judgments (rather than beliefs about cues), three of the four available visual cues were associated with judgments of deception, and none of these patterned with their accuracy as actual cues. In Zuckerman, DePaulo, and Rosenthal (1981), participants' beliefs about gestural cues all tended towards the same direction: More movement signals dishonesty. Even speakers' post-hoc perceptions of their own gestures when lying have been found to be at odds with how they actually behaved: Vrij, Semin, and Bull (1996) found that after partaking in two interviews — one in which they were truthful, the other dishonest — participants believed that their movements increased when lying, even though a decrease actually occurred. Furthermore, whether or not participants were informed before-hand that deception is usually associated with a decrease in movements had no effect on subsequent beliefs about their own behaviour.

The divergence between actual and perceived cues to deception patterns with the consistent finding that we are bad at discriminating lies from truths. In reviewing 39 studies in the detection of deception, Vrij et al. (2000) found an average accuracy rate of 56.6% — only just above chance. One suggested reason for this poor rate of detection was that listeners often look for the wrong cues in attempting to detect deceit (as evidenced in DePaulo, Rosenthal, Rosenkrantz, and Green 1982), perhaps resulting from being unaware of their own behaviour during lying (as in Vrij et al. 1996), combined with an inaccurate model of what deceitful behaviour looks and sounds like. This inaccurate model may be predicated on introspection as a speaker rather than experience as a listener: Speakers believe themselves to be more disfluent (Zuckerman, Koestner, & Driver, 1981) as well as to produce more movement (Vrij et al., 1996) when lying.

In order to better understand how speech cues influence the judgments listeners make about speakers' honesty, recent research by Loy et al. (2017) investigated the time-course of the disfluency-lying bias. Loy et al. (2017) used a visual world paradigm in which participants were presented with two objects along with utterances describing the location of some treasure purportedly hidden behind one of the objects. These utterances were presented as having been elicited in a previous experiment, and could be either truthful or dishonest. Crucially, Loy et al. (2017) manipulated the manner of spoken delivery, with half of the experimental items containing a speech disfluency. Participants were tasked with choosing where they *believed* the treasure to really be hidden — the object named in the utterance (indicating a judgment of honesty), or a distractor (indicating dishonesty). Participants were more likely to judge disfluent utterances as dishonest than fluent ones (as indicated by more clicks on the distractor on disfluent trials). Importantly, their results showed an early bias in both eye and mouse movements towards the non-referent object, providing evidence that listeners are capable of accessing and integrating contextual information (manner of spoken delivery) to alter their pragmatic interpretation of an utterance

Little is known, however, about whether such rapid integration of cues might extend to the visual channel of communication. To date, research into how speech and gesture interact in language comprehension has focused largely on how representational gestures can influence the literal interpretation of the message. Gestures which mismatch accompanying speech (e.g., "square" while gesturing a triangle) have been found to interfere with a listener's on-line processing, suggesting that speech and gesture are integrated in real-time to form a unified system in comprehension (see Habets, Kita, Shao, Özyurek, and Hagoort 2011; Kelly, Creigh, and Bartolotti 2010). While evidence supports the on-line integration of representational gestures with speech, little research has investigated how listeners process the 'unintentional' movements made by speakers. Furthermore, the interplay between gesture and pragmatic comprehension also remains largely under-studied. Representational gesturing has been found to influence listeners' pragmatic interpretations, such as the referents of indirect requests (see Kelly and Barr 1999), although measurement has been limited to off-line comprehension. In the field of deception research, the link between gestures and perceived deceit has likewise been studied only in terms of after-the-fact judgments, or assessing listeners' explicit beliefs about cue validity (see Vrij and Semin 1996; Zuckerman, Koestner, and Driver 1981) Less is known about the time course with which listeners integrate such information in the visual modality to form pragmatic judgements of deception.

The present experiments investigate how listeners process the unintentional movements made by a speaker, specifically how these inform judgments about a speaker's (dis)honesty. We extend the 'treasure game' paradigm from Loy et al. (2017) to include a video of a (potentially deceptive) speaker describing the location of some hidden treasure on the screen while listeners attempt to guess the location of the treasure based on whether they believe the speaker to be lying or telling the truth. Crucially, we manipulate the presence or absence of gestures made by the speaker in the video. Experiment 1 assesses which types of gestures listeners might associate with deception. In Experiment 2 we focus on if, how, and when, adaptor gestures (fidgeting movements) influence listeners' judgments of deception. Additionally, we ask whether listeners' judgments are predicted by their perceptions of how nervous they think the speaker to be in each video.

Experiment 1

Experiment 1 makes use of eye- and mouse-tracking to investigate the time course of listeners' judgments about the honesty of an utterance, and how these judgments are influenced by the occurrence of various types of gestures. The experiment was presented as a 'lie detection game', with each trial presenting a video of a potentially deceptive speaker describing the location of some hidden treasure on the screen (behind one of two possble objects). Participants are tasked with clicking on where they believe the treasure to be hidden based on their judgement of the speaker's honesty. We employ 3 types of gestures—trunk movements, adaptors, and changes in posture—and investigate whether listeners associate any of these with deception.

Materials

Visual stimuli consisted of the same 120 line drawings from Snodgrass and Vanderwart (1980) which were used in Loy et al. (2017), presented in pairs across sixty trials. In each trial, the speaker named one object (referent) as that which concealed the treasure; the other object is hereafter named the distractor. Each referent was associated with a recording specifying the image as the object that the treasure was hidden behind ("The treasure is behind the <referent>"). Along with these images, each trial contained a video of a person who was purported to be the speaker of the utterances. So that videos could be counterbalanced across referents, and thus across utterances, the face of the person in the video was blurred. This meant that, when presented with a given utterance, it was believable that both audio and visual stimuli had been produced concurrently.

Sixty videos (30 Gesture, 30 No-gesture) were created. The thirty no-gesture videos were comprised of ten recordings (each presented thrice) of a speaker sitting motionless with her hands either side of a tablet on a table, upon which the location of treasure was purported to be displayed. Ten videos presented the speaker making one of five trunk movements, ten presented various adaptor gestures (e.g. finger tapping, head tilts), and ten presented the speaker sitting motionless but in a different position to that of the no-gesture videos (e.g. hand on chin, arms crossed, etc).

In order to make the stimuli look as natural as possible, a variable video-to-speech-onset was adopted for the 30 gesture videos. For trunk movements, the frame at which the movement ended was identified and used as the point of utterance onset (Mean onset = 1430ms, SD = 440ms). This allowed the duration of the trunk movement to be interpretable as speech initiation time, which could in turn potentially be associated with the speaker preparing a dishonest utterance. To control for any effect of gesture being explained as simply a sensitivity to the duration of pre-utterance video, these durations were matched in the no-gesture videos. For adaptor movements, the gestures were allowed to overlap with speech in order to make the audio and video more believable, with the duration of overlap determined individually for each video (Mean = 1370ms, SD = 410ms). These durations were matched in the videos that presented the speaker sitting in a different posture to the no-gesture videos.

As with (Loy et al., 2017), 20 critical referents were counterbalanced across two lists, each containing 10 gesture and 10 no-gesture videos. The 10 gesture videos were trunk movements since we predicted these to be the most likely to elicit judgements of deception associated with a gesture (Vrij & Semin, 1996) The remaining 40 referents were randomly paired with one of the remaining videos (10 adaptors, 10 different postures, 20 no-gesture) for each participant, with no repetition of referents across videos.

Procedure

Stimuli were displayed on a 21 in. CRT monitor, placed 850 mm from an Eyelink 1000 Tower-mounted eye-tracker which tracked eye movements at 500 Hz (right eye only). Audio was presented in stereo from speakers on either side of the monitor. Mouse coordinates were sampled at the frame rate of the videos (25 fps). The experiment was presented using OpenSesame version 3.1 (Mathôt, Schreij, & Theeuwes, 2012). Eye movements, mouse coordinates and object clicked (referent or distractor) were recorded for each trial.

Figure 1 presents a sample trial from the experiment. Between trials, participants underwent a manual drift correct, after which the fixation dot turned red for 500ms. After this, the two objects (referent and distractor) were displayed on the screen for 2000ms. The video then appeared and the cursor was centred and made visible. Playback of the utterance began after the variable speech onset associated with each video (Mean = 1410ms, SD = 410 ms).

The instructions emphasised that the videos participants saw were recorded from a previous experiment, in which the speaker had to describe the location of some hidden treasure with the aim of misleading the listener into choosing the wrong location. Participants were instructed to click on the object behind which *they believed* the treasure to be hidden, with the overall aim of accumulating as much treasure as they could across the experiment. Participants received no feedback after their object clicks, except on bonus trials, which are described in the next section.

Participants completed five practice trials (one of which was presented as a bonus round) prior to the main experiment. Two of these were static, two displayed the speaker in different postures, and one displayed the speaker making a trunk movement.

Bonus Rounds

To maintain motivation throughout the study, participants were told that there were a number of "hidden bonus rounds" which offered more treasure than regular rounds. 25% of trials (half in the gesture condition; half in the no-gesture condition) were randomly designated as bonus rounds for each participant. These trials were visually identical to regular trials, with the exception of a message informing participants that they had successfully located bonus treasure following their mouse click (regardless of the object chosen).

Participants were also told that the top scorers would be able to enter their names on a high-score table, which was shown at the beginning of the experiment.

Post-test Questionnaire

Participants were asked to complete a short post-test questionnaire. The questionnaires contained three questions, the most important of which asked if participants noticed anything odd about the visual or audio stimuli. Any participant who indicated that they noticed anything unusual was then verbally questioned, to decide whether they believed that the speech and gesture had been produced naturally and simultaneously. All participants were subsequently debriefed (told that the audio and video were created separately and stitched together), and asked again verbally if they noticed anything unusual in that respect.

Results

Twenty-four native English speaking participants took part in the experiment for a desired sample size of twenty. Participants were recruited from the University of Edinburgh community, and participated in return for a payment of £4. Data from four participants who indicated suspicion of the proposed origins of the audiovisual stimuli based on the post-test questionnaire were removed from all analyses.

Analysis

Analysis was carried out in R version 3.4.4 (R Core Team, 2017), using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Trials in which participants did not click on either the referent or distractor (0.003% of trials) were excluded from all analyses.

Object clicked (referent or distractor) was modeled using mixed effects logistic regression, with fixed effects of gesture type (No Gesture, Different Postures, Trunk Movements, Adaptor Gestures) and video-to-speech duration (Z-scored), thus controlling for any possible effect of perceived speech latency on judgments of dishonesty. Random intercepts and slopes for gesture type and video-to-speech were included by-participant, along with random intercepts by-referent. Reaction times (measured from referent onset) were modelled with the same fixed and random effect structure. Following Lo and Andrews (2015), we compared mixed effects logistic regression models which specified an identity link function, assuming gaussian, gamma and inverse gaussian distributions.

In previous studies using the treasure game paradigm, eye- and mouse- movements have been analysed over the time window starting at the onset of the referent name and extending for 800 ms; just beyond the duration of the longest referent (776 ms). Because the current study includes in the analysis all available referents (rather than the subset used in previous experiments), this window is extended to 0–1100 ms to include the duration of the longest referent (1062 ms).

Eye fixation data was averaged into 20 ms bins (of 10 samples) prior to analysis. For each bin, we calculated the proportions of time spent fixating referent or the distractor, resulting in a measure of the proportions of fixations on either object over time.

The position of the mouse was sampled every 40 ms. Using the X coordinates only, we calculated the number of screen pixels moved and the direction of movement (towards either referent or distractor). We then calculated the cumulative distance travelled towards each object over time as a proportion of the cumulative distance travelled in both directions up until that time bin. Movements beyond the outer edge of either object were considered to be 'overshooting' and were not included in calculations (0.8% of samples). Eye- and mouse- biases were calculated from the proportions of referent to distractor fixations, and were subsequently empirical logit transformed (Barr, 2008). In these measures, a value of zero indicates no bias towards either object, and positive and negative values indicate a bias towards the referent and distractor respectively.

Eye and mouse data was modelled over the time window from 0 to 1100 ms post-referent onset using linear mixed effects models, with fixed effects of time, gesture type, and their interaction. Random intercepts and slopes for time were included both by-referent and by-participant, along with by-participant random effects of gesture type. Following Baayen (2008), we considered effects in these models to be significant where $|t| > 2$.

As visual inspection of the time-course of fixations towards either object suggested that there was a later effect of gestures in participants' decision of which object to click on, growth curve analysis (See Mirman, Dixon, and Magnuson 2008) was used to investigate this further. This analysis was conducted over the time window from 0 to 1815 ms post-referent onset included 3 degrees of orthogonal polynomials for time, based on the pattern of the elogit referent-distractor bias having 2 turning points. These time polynomials, along with their interaction with gesture type, were included as fixed effects in a linear mixed effects model, with random intercepts and slopes for all time polynomials both by-participant and by-referent.

Object clicks

Across the experiment, participants clicked on the referent in 55% of trials and the distractor in only 45%. Table 1 shows the percentage of clicks across all participants to either object following different types of gesture. When presented with an utterance accompanied by no-gesturing, participants showed a bias toward a final interpretation of the utterance as truthful, with more clicks to the referent than the distractor $\beta = 0.62$, SE = 0.16, $p < 0.001$. Analysis by individual gesture types showed that all types saw a reduction in this bias, with adaptor gestures ($\beta = -1.03$, SE = 0.34, $p < 0.005$) showing a greater change than different postures and trunk movements ($\beta = -0.72$, SE = 0.31, $p < 0.05$ and $\beta = -0.62$, SE = 0.26, $p < 0.05$ respectively). The duration of video shown prior to the beginning of speech was not found to be associated with which object was eventually clicked.

Comparisons — via both AIC and BIC — of reaction time models suggested that an inverse gaussian distribution provided the best fit to the observed data. Neither gesture type nor duration of video prior to speech was associated with a significant change in reaction times.

Eye movements

Figure 2 shows the time-course of fixations to referents and distractors over 2000 ms from referent onset, split by each type of video.

Analyses conducted over the period from referent onset to 1100 ms post-onset (duration of the longest referent) showed that, when presented with a no-gesture video, participants displayed a fixation bias towards the referent which increased over time

($\beta = 1.15$, SE $= 0.29$, $t > 2$). For videos which presented the speaker either in a different posture, or producing an adaptor gesture, this increasing bias to the referent was significantly reduced ($\beta = -0.97$, SE $= 0.13$, $t > 2$, $\beta = -0.59$, SE $= 0.13$, $t > 2$), but this was not the case for videos of trunk movements ($\beta = -0.25$, SE $= 0.15$, $t = 1.67$).

Growth curve analysis over the period from referent onset to 1815 ms post-onset (mean click time) showed that when presented with a no-gesture video, participants showed an tendency to fixate on the referent over the course of this time period ($\beta = 0.67$, SE $= 0.11$, $t > 2$). Significant effects of gesture type on the intercept term indicated lower overall referent fixations during this period after viewing a different posture ($\beta = -0.36$, SE $= 0.04$, $t > 2$), a trunk movement ($\beta = -0.29$, SE $= 0.4$, $t > 2$), or an adaptor gesture ($\beta = -0.67$, SE $= 0.4$, $t > 2$), in comparison to the no gesture videos. Significant interactions of gesture type and the linear time coefficient indicate that this reduction in referent-bias dependent on gesture type increases over the course of the window, with adaptor gestures showing a larger reduction relative to no gesture videos ($\beta = -116.89$, SE $= 11.62$, $t > 2$) than different postures ($\beta = -75.41$, SE $= 11.65$, $t > 2$) and trunk movements ($\beta = -58.78$, SE $= 12.97$, $t > 2$). Figure 3 shows the time-course of the elogit referent-distractor bias, alongside the fitted values from the growth curve model.

Mouse movements

Figure 4 shows the time-course of the proportions of cumulative distance the mouse moved towards the referent and distractor for 2000 ms from referent onset, split by each type of video. Analysis on the time window from 0 to 1100 ms post-referent onset patterned with the eye-tracking data: When the speaker made no gesture, participants showed an increasing tendency to move towards the referent over this period ($\beta = 1.08$, SE $= 0.22$, $t > 2$). As with the eye-movements, different postures and adaptor gestures resulted in a weakening of this referent-bias ($\beta = -0.81$, SE $= 0.13$, $t > 2$ and $\beta = -0.77$, SE $= 0.13$, $t > 2$ respectively), but trunk movements did not ($\beta = -0.10$, SE $= 0.14$, $t = 0.69$).

Discussion

Experiment 1 investigated how the pragmatic inferences listeners make about a speaker's honesty are influenced by the presence of different types of movements and postures, and which of these provide the best means of studying the time-course of such inferences. As in previous studies using this paradigm (King, Loy, & Corley, 2018; Loy et al., 2017), participants showed a tendency to interpret an utterance as truthful when it had been presented without any potential cue to deceit. Utterances presented alongside different postures and movements weakened this tendency, but only the presence of adaptor gesturing resulted in significantly more judgments of deception than truthfulness (indicated by clicks to the distractor over the referent).

The influence of the visual channel was also evident in participants' early stages of utterance processing. Across both gesture and no-gesture videos, participants showed an initial tendency to fixate and move the mouse toward the referent over distractor; however, videos of different postures and adaptor gestures resulted in a reduction of this tendency. Interestingly, this early effect was not found for videos of trunk movements. One possible explanation for this could be due to the fact that in all videos of trunk movements, the speaker completed the gesture prior to the onset of speech. This meant that at the point of referent onset, the audiovisual information immediately available to the listener was comparable to the no gesture videos. Similarities may be drawn here to the subtle differences when presented with either utterance-initial or utterance-medial speech disfluencies in Loy et al. (2017): The temporal proximity of a cue may potentially influence how much it reduces the initial referent-bias.

When presented with gestures, the point at which listeners' preference toward the referent tended to be matched (or overtaken, in the case of adaptor gestures) by a preference for the distractor appeared at approximately 1000 ms post-referent onset. Contrastingly, participants in Loy et al. showed little to no initial referent-bias following a disfluent utterance, with a distractor-bias appearing approximately 600 ms post-referent onset following both utterance-initial and utterance-medial disfluencies.

There are several possible explanations for the later effect found in the current study. Firstly, the inclusion of longer referents, along with the fact that not all referent-distractor pairs were controlled for phoneme onset, may have resulted in establishing reference being delayed. Secondly, the inclusion of videos in the experimental stimuli may have detracted from participants' fixations to the two objects. This could also explain the differences in the early window (0 to 1100 ms) between the types of gesturing: The reduction in the referent-bias may be a result of the video being more salient in both the different posture and the adaptor gesture conditions, and not evidence of the emerging gesture-lying bias. That said, however, the initial referent-bias appears at approximately 300 ms post-referent onset in all conditions — a point comparable to previous studies with this paradigm (King et al. 2018; Loy et al. 2017).

Lastly, it might be that the integration of gestural cues during judgments of deception is simply slower than it is for speech cues, taking longer to counteract the initial tendency to look at a named object. This is perhaps because the link from gestures to perceived deception may be subject to different processing mechanisms than the link from disfluency to deception. Loy et al. (2017) showed that listeners' disfluency-lying bias occurs very early in comprehension, suggesting either a reliance upon a rule-of-thumb assocation between the two, or extremely fast inferential processing which links disfluency to deception via, for example, cognitive effort involved in planning a dishonest utterance (see King et al. 2018). However, this might not be the case for gestures. Speech disfluency can be attributed directly to the processes of speech production: This allows listeners to draw a straightforward inferential link from the way a given utterance is delivered to the content of that utterance, be it pragmatic (e.g. lying) or literal (e.g. naming difficulty, Arnold, Kam, and Tanenhaus cf. 2007). Co-speech movements, however, can be attributed to many things, some only indirectly related to the production of speech, and some completely unrelated (e.g. scratching an itch). Whilst there is evidence to suggest that cognitive effort increases gesturing (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), an alternative intuitive line of reasoning for a listener would, for instance, link gestures to nervousness/anxiety, and link nervousness to deception. Indication of this can be found in participants' responses to the post-test questionnaire. Four participants felt that they had made their judgements based on whether the speaker looked tense or relaxed (with 1 even mentioning that videos in which the speaker moved looked more relaxed, and so signalled honesty). In fact, only 2 participants explicitly linked an *increase* in movement with lying; 9 mentioned using body language and gestures to make their judgments but specified neither rationale nor direction; and 15 described perceiving differences in spoken delivery (such as in the speaker's intonation or hesitations) which they believed signalled lying. It might be that participants made their decisions not directly based on the presence or absence of a gesture, but on something which took longer to establish, such as how nervous the speaker looked over the course of the video.

The results from Experiment 1 support previous findings in the deception literature that listeners' judgments of deception are influenced by changes in the speaker's postures and body language. However, it is possible that the time-course of these judgments was made less clear by stimuli which did not capture the underlying mechanism by which listeners associate gesture with lying. To clarify this, we conducted Experiment 2, a simplified version of the experiment in which we focused on one type of gesturing. Additionally, we included a task after the eye-tracking segment asking participants to rate each video for how nervous they perceived the speaker to be.

Experiment 2

Experiment 2 focused on the influence of adaptor gestures on judgments of dishonesty, and looked at whether these judgments paralleled with perceptions of nervousness in the speaker. We focused on adaptor gestures since these were found to have the largest effect on listeners' judgments of dishonesty in Experiment 1. If gesture is linked to deception via perceived nervousness, then gestures which are rated as more nervous should result in more judgements of dishonesty. Using the same paradigm as Experiment 1, participants in Experiment 2 heard utterances accompanied by a video of a speaker either producing an adaptor gesture or no gesture, and were tasked with making an implicit judgement on whether the speaker was lying or telling the truth. After the task, participants were asked to rate each video (without audio) on how nervous the speaker looked. This provided us with a measure of association between the speaker's body language and their perceived nervousness.

Materials

A subset of 40 images (20 referents; 20 distractors) from those in Experiment 1 were used across twenty trials. As in Experiment 1, these images were displayed in referent-distractor pairs, with each pair shown alongside a recorded utterance naming the referent as the location of the treasure. Following Loy et al. (2017), we used the same set of referents and distractors which had been matched for both ease of naming and familiarity.

As in Experiment 1, each pair of images and recorded utterance was presented alongside a video clip of a person purported to be the speaker of the utterance. Twenty video clips (10 adaptors; 10 no-gesture) were used. Based on participant feedback from Experiment 1, care was taken to ensure that the No Gesture videos presented the speaker in a relaxed posture. Adaptor gestures were based on descriptions of anxious non-verbal behaviour from Gregersen (2005). Videos were chosen from a pre-test which asked participants to rate 28 silent videos (18 different adaptor gestures; 10 no-gestures) for their perceived nervousness of the speaker. 10 native english speakers were told that they were going to watch videos (without audio) of someone being questioned in a stressful situation, and were asked to rate how nervous the speaker looked in each video (1: very relaxed, 7: very nervous). The 10 adaptor gestures with the highest ratings (Mean = 4.1, SD = 1.5) were included in the experiment, along with the 10 no-gesture videos (Mean = 1.9, SD = 1.1).

The 20 referents were counterbalanced across two lists such that each referent that occurred with a gesture video in the first list occurred with a no-gesture in the second. The pairings of referents with specific videos/gestures within each condition was randomised on each run of the experiment.

Procedure

The experiment procedure was identical to that of Experiment 1 with two minor changes. First, speech initiation time (the duration between video playback and utterance onset) was set a fixed constant of 1170 ms after the beginning of the video on each trial. Second, since the experiment was considerably shorter, we did not include any 'bonus' trials which displayed a message stating that treasure had been found after an object click. Hence, participants did not receive any feedback in any of the trials in this experiment.

After the main task, participants were asked to watch all 20 videos again, without audio, and asked to rate how nervous they thought the speaker looked (using the same 1-7 scale as described above). Participants then completed the same post-test questionnaire as in Experiment 1, with data being excluded from analysis based on the same criteria.

Results

Twenty-three native English speaking participants took part in exchange for £3 compensation. Data from three participants were excluded due to suspicion of the audiovisual stimuli being scripted (based on the post-test questionnaire and questioning during debrief), hence the final dataset included data from 20 participants.

Analysis

We followed the same analysis strategy as was used for Experiment 1, with the experimental manipulation of gesture being a dichotomous Gesture vs. No Gesture. Trials which did not result in a click to either object (0.8%) were excluded from analyses.

Analyses of object clicks and reaction times did not control for video-to-speech duration, since this was controlled in the experimental design. To investigate whether listeners' deception judgments were predicted by their subsequent ratings of perceived nervousness, analysis of object clicks included participants' post-test ratings (Z scored) of each video as a fixed effect, along with the interaction with gesture.

The time window of analysis for eye- and mouse- movements was reduced to the 800 ms following referent onset, based on the fact that the subset of referents included in this experiment had a maximum duration of 776 ms. For the mouse movement analysis, movements beyond the outer edge of either object were excluded from analyses (1% of samples). Because Experiment 2 fully counterbalanced gesture across all referents, random effects of gesture and time and their interaction were included both by-referent and by-participant.

Object clicks

Across the experiment, participants clicked on the referent in 53% of trials and the distractor in 47%. Table 2 shows the proportions of clicks to either object following videos displaying either an adaptor gesture or no gesture. As in Experiment 1, participants showed a bias toward a final interpretation of the utterances without gestures as truthful, with more clicks to the referent than the distractor $\beta = 1.15$, SE = 0.37, $p < 0.005$. Patterning with the adaptor gestures in Experiment 1, utterances presented with an adaptor gesture here resulted in bias towards participants clicking the distractor ($\beta = -2.45$, SE = 0.51, $p < 0.001$). Inclusion of participants post-test ratings of how nervous they perceived the speaker to be in each video did not improve model fit (χ^2=0.46) There was no effect of gesture on time to click.

Eye movements

Figure 5 shows the time-course of fixations to referents and distractors over 2000 ms from referent onset, split by presence of gesture. Analyses conducted over the period from referent onset to 800 ms post-onset patterned with results from Experiment 1: When presented with an utterance unaccompanied by a gesture, participants displayed a fixation bias towards the referent which increased over time ($\beta = 3.03$, SE = 0.77, $t > 2$). When presented with an utterance accompanied by an adaptor gesture, this bias was greatly reduced ($\beta = -3.00$, SE = 1.02, $t > 2$).

Mouse movements

Figure 6 shows the time-course of the proportions of cumulative distance the mouse moved towards the referent and distractor for 2000 ms from referent onset, split by presence of gesture. Analysis on the period from 0 to 800 ms post-referent onset patterned with the eye-tracking data: Following no gesture videos, participants showed a tendency to move the mouse increasingly towards referent over this period ($\beta = 1.90$, SE = 0.39, $t > 2$). As with eye-movements, this referent-bias was greatly reduced following an adaptor gesture ($\beta = -2.46$, SE = 0.77, $t > 2$)

General discussion

Across both experiments, participants showed a slight overall tendency to interpret an utterance as honest rather than dishonest, although this only differed from chance in Experiment 1. This bias is in line with results from previous research on deception, which shows that listeners have a general tendency to judge a speaker as truthful (Vrij et al. 2000). Our results show that the body language and gestures which accompany a spoken utterance affects the judgments which listeners make about the utterance's veracity. Utterances presented with the speaker in a neutral posture and not gesturing bias listeners towards believing the speaker to be truthful, as shown by increased tendency to fixate on, and move the mouse towards, the object which was named by the speaker. This result parallels the finding in Loy et al. (2017) where fluent (as opposed to disfluent) utterances saw a bias to infer the speaker to be truthful, and suggests that listeners may have an implicit honesty-bias when faced with no obvious potential cue to deception.

Consistent with previous research in deception judgments and beliefs about cues to deception, participants linked an increase in movement with dishonesty (Zuckerman, DePaulo, & Rosenthal, 1981). Interestingly, participant's subsequent ratings of how nervous the speaker looked in each video did not pattern with their final judgments beyond whether or not the video presented a gesture, suggesting that listeners' associations between gesture and lying is not necessarily a consequence of perceived anxiety in the speaker. Although the mechanism behind listeners' associations between gestures and deception remain unclear, our time course results show that gestural cues influence pragmatic judgments from the early stages of comprehension. Extending previous work demonstrating a link between increased gesturing and perceived dishonesty, we show that these pragmatic judgments which listeners made began during the on-line processing of speech, becoming apparent alongside the unfolding of the critical linguistic and visual input. The speed with which these judgments were made — especially in Experiment 2 — suggest that visual information can modulate listeners' judgments about a speaker's intentions concurrently with the integration of lexical information. Our results show that the integration of the visual channel can have a rapid and direct effect on pragmatic judgements, supporting the idea that communication is fundamentally multimodal: Speech and gesture interactively codetermine meaning.

Different types of gestures and postures influence deception judgments to different degrees, affecting the time-course of these judgments accordingly. In Experiment 1, a reduction in the tendency of eye- and mouse- movements towards the referred-to object was apparent when the speaker was either in a different posture or made a trunk movement prior to speaking, but there was no bias towards eventually interpreting utterances accompanied by these types of gesture as either truthful or dishonest. Utterances accompanied by adaptor gestures, however, were associated with more judgments of dishonesty than of truthfulness, and this patterned with a greater reduction in the early eye- and mouse- movement biases towards the referent. Experiment 2 provided a clearer account of the influence of adaptor gestures on deception judgments at the early moments of referent comprehension, with the tendency to look — and move the mouse — towards the referent being largely attenuated when the utterance was presented with a gesture. However, following adaptor gestures in both experiments, the bias towards the distractor — signifying perceived dishonesty — over the referent appeared approximately 1000 ms after the referent began, at a later point than previous versions of this paradigm in which speech disfluencies were found to modulate judgements of deception This may reflect differences in how the comprehension system integrates auditory information, such as a disfluency, and visual information, such as a gesture. In particular, the integration of these cues to inform deception judgments might be more gradual than the integration of spoken cues. To better understand how cues to deception in different modalities affect comprehension, further research would require investigating the effect of disfluency when the visual channel is also available — for example, studying the time course of deception judgments when faced with one or both of a disfluency and an adaptor gesture.

References

Arciuli, J., Villar, G., & Mallard, D. (2009). Lies , lies and more lies. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2329–2334).

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930. Retrieved from `http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.33.5.914` doi: 10.1037/0278-7393.33.5.914

Baayen, R. (2008, dec). Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. *Sociolinguistic Studies*, *2*(3), 353. Retrieved from `http://www.equinoxjournals.com/ojs/index.php/SS/article/view/5557` doi: 10.1558/sols.v2i3.471

Barr, D. J. (2008, nov). Analyzing âĂŸvisual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0749596X07001015` doi: 10.1016/j.jml.2007.09.002

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015, oct). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1215–1225. Retrieved from `http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdfhttp://journals.sagepub.com/doi/10.1177/009286150103500418http://www.jstatsoft.org/v67/i01/` doi: 10.18637/jss.v067.i01

Benus, S., Enos, F., Hirschberg, J., & Shriberg, E. (2006, may). Pauses in Deceptive Speech. In *Speech prosody.*

Cohen, D., Beattie, G., & Shovelton, H. (2010). Nonverbal indicators of deception: How iconic gestures reveal thoughts that cannot be suppressed. *Semiotica*, *2010*(182), 133–174. doi: 10.1515/semi.2010.055

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*(2), 203–243. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.111.2.203 doi: 10.1037/0033-2909.111.2.203

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.70.5.979 doi: 10.1037/0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–112. Retrieved from http://doi.apa.org/getdoi.cfm?doi=10.1037//0033-2909.129.1.74 doi: 10.1037//0033-2909.129.1.74

DePaulo, B. M., Rosenthal, R., Rosenkrantz, J., & Green, C. R. (1982, dec). Actual and Perceived Cues to Deception: A Closer Look at Speech. *Basic and Applied Social Psychology*, *3*(4), 291–312. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15324834basp0304{_}6 doi: 10.1207/s15324834basp0304_6

Ekman, P. (1989). Why Lies Fail and What Behaviors Betray a Lie. In *Credibility assessment* (pp. 71–81). Dordrecht: Springer Netherlands. Retrieved from http://link.springer.com/10.1007/978-94-015-7856-1{_}4 doi: 10.1007/978-94-015-7856-1_4

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychological science : a journal of the American Psychological Society / APS*, *12*(6), 516–522. doi: 10.1111/1467-9280.00395

Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, *38*(3), 388–400. doi: 10.1111/j.1944-9720.2005.tb02225.x

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011, aug). The Role of Synchrony and Ambiguity in SpeechâĂŞGesture Integration during Comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854. Retrieved from http://www.mitpressjournals.org/doi/10.1162/jocn.2010.21462 doi: 10.1162/jocn.2010.21462

Kelly, S. D., & Barr, D. J. (1999). Offering a Hand to Pragmatic Understanding : The Role of Speech and Gesture in Comprehension and Memory. , *592*, 577–592.

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010, apr). Integrating Speech and Iconic Gestures in a Stroop-like Task: Evidence for Automatic Processing. *Journal of Cognitive Neuroscience*, *22*(4), 683–694. Retrieved from http://www.mitpressjournals.org/doi/10.1162/jocn.2009.21254 doi: 10.1162/jocn.2009.21254

King, J. P., Loy, J. E., & Corley, M. (2018). Contextual Effects on Online Pragmatic Inferences of Deception. *Discourse Processes*, *55*(2), 123–135. Retrieved from https://doi.org/10.1080/0163853X.2017.1330041 doi: 10.1080/0163853X.2017.1330041

Lo, S., & Andrews, S. (2015, aug). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*(August), 1–16. Retrieved from http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.01171/abstract doi: 10.3389/fpsyg.2015.01171

Loy, J. E., Rohde, H., & Corley, M. (2017, may). Effects of Disfluency in Online Interpretation of Deception. *Cognitive Science*, *41*, 1434–1456. Retrieved from http://doi.wiley.com/10.1111/cogs.12378 doi: 10.1111/cogs.12378

Mathôt, S., Schreij, D., & Theeuwes, J. (2012, jun). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. Retrieved from http://www.springerlink.com/index/10.3758/s13428-011-0168-7 doi: 10.3758/s13428-011-0168-7

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008, nov). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0749596X07001313 doi:

10.1016/j.jml.2007.11.006

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

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. Retrieved from `http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.6.2.174` doi: 10.1037/0278-7393.6.2.174

Vrij, A. (1995, jan). Behavioral Correlates of Deception in a Simulated Police Interview. *The Journal of Psychology*, *129*(1), 15–28. Retrieved from `http://www.tandfonline.com/doi/abs/10.1080/00223980.1995.9914944` doi: 10.1080/00223980.1995.9914944

Vrij, A., Kneller, W., & Mann, S. (2000, feb). The effect of informing liars about Criteria-Based Content Analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology*, *5*(1), 57–70. Retrieved from `http://doi.wiley.com/10.1348/135532500167976` doi: 10.1348/135532500167976

Vrij, A., & Semin, G. R. (1996, mar). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior*, *20*(1), 65–80. Retrieved from `http://link.springer.com/10.1007/BF02248715` doi: 10.1007/BF02248715

Vrij, A., Semin, G. R., & Bull, R. (1996, jun). Insight Into Behavior Displayed During Deception. *Human Communication Research*, *22*(4), 544–562. Retrieved from `https://academic.oup.com/hcr/article/22/4/544-562/4564904` doi: 10.1111/j.1468-2958.1996.tb00378.x

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and Nonverbal Communication of Deception. In (pp. 1–59). Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S006526010860369X` doi: 10.1016/S0065-2601(08)60369-X

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, *6*(2), 105–114. Retrieved from `http://link.springer.com/10.1007/BF00987286` doi: 10.1007/BF00987286

Table 1

Breakdown of mouse clicks recorded on each object (referent or distractor) by type of gesture for Experiment 1

	No Gesture	Different Posture	Trunk Movement	Adaptor Gesture
Clicks to Referent	63.8%	48.0%	49.7%	41.5%
Clicks to Distractor	36.2%	52.0%	50.3%	58.5%

Table 2

Breakdown of mouse clicks recorded on each object (referent or distractor) by presence of gesture for Experiment 2

	No Gesture	Gesture
Clicks to Referent	80.9%	24.2%
Clicks to Distractor	19.1%	75.8%

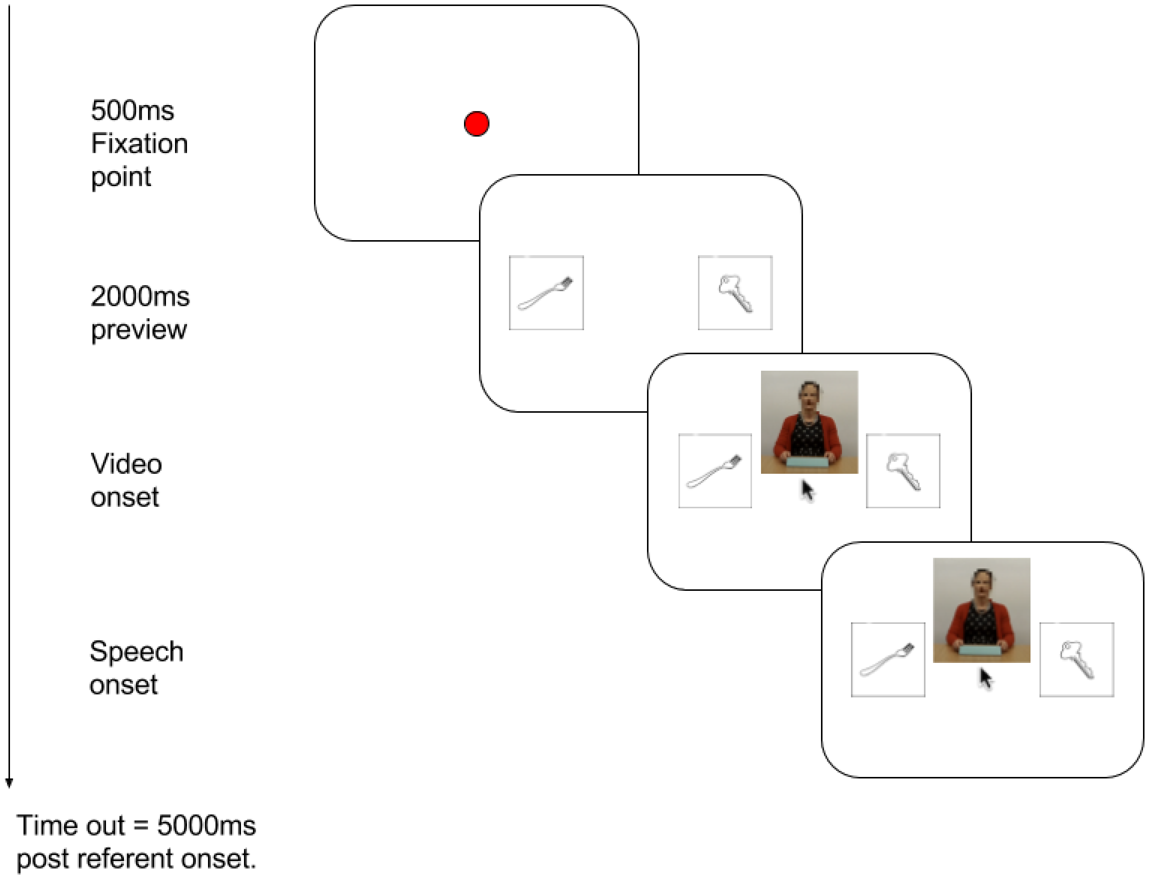

Figure 1. Procedure of a given trial, Experiment 1

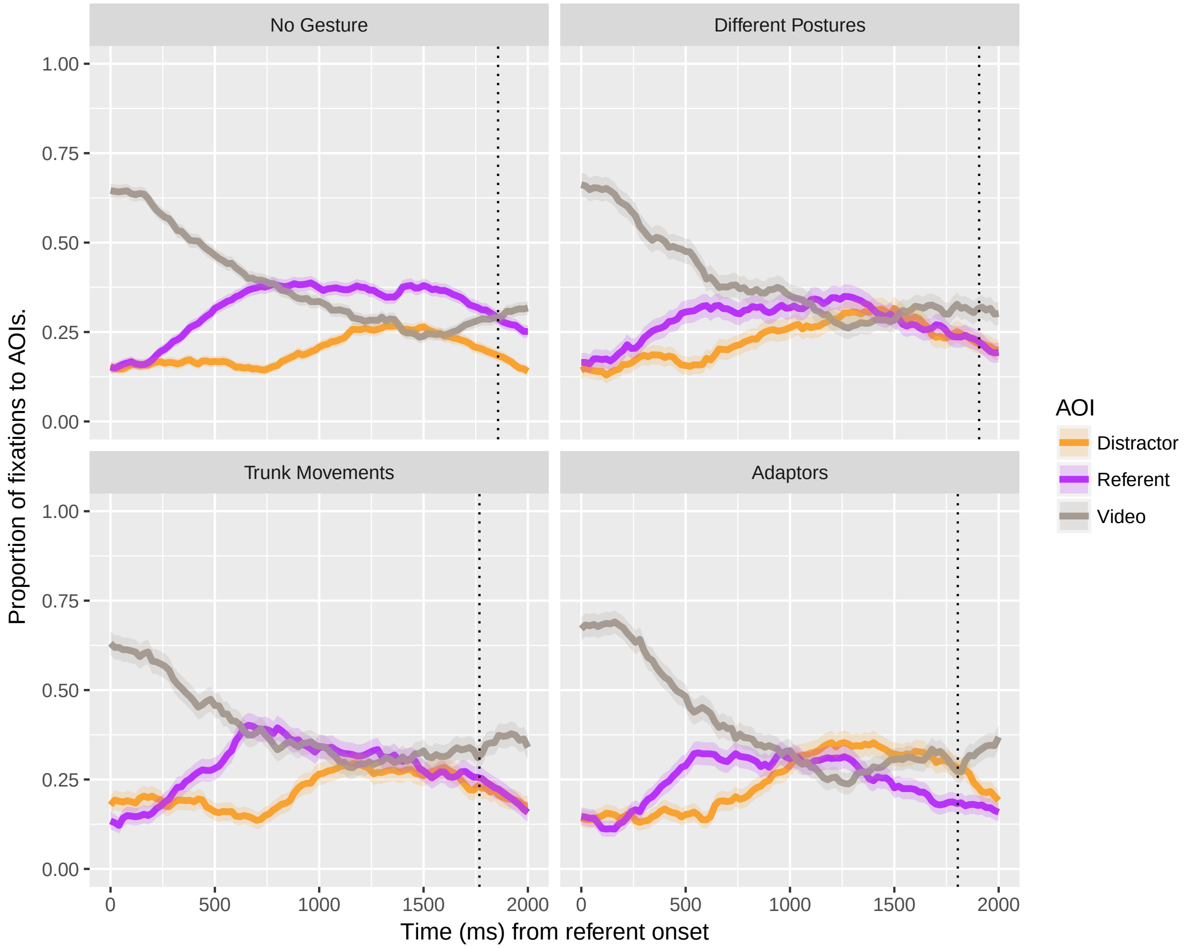

Figure 2. Eye-tracking results for Experiment 1: Proportion of fixations to each object (referent, distractor, video), from 0 to 2000 ms post-referent onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ± 1 standard error of the mean.

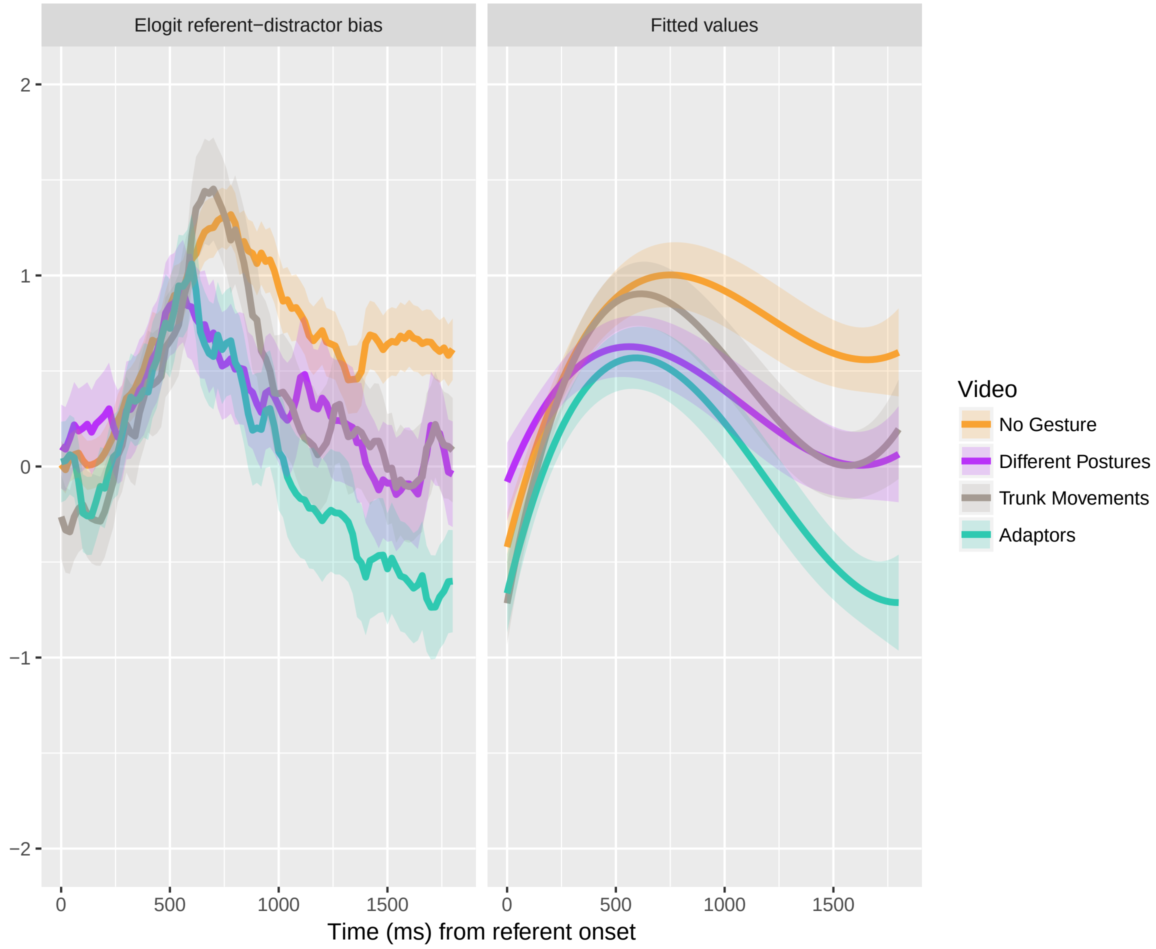

Figure 3. Elogit referent-distractor bias, and fitted values from growth curve analysis

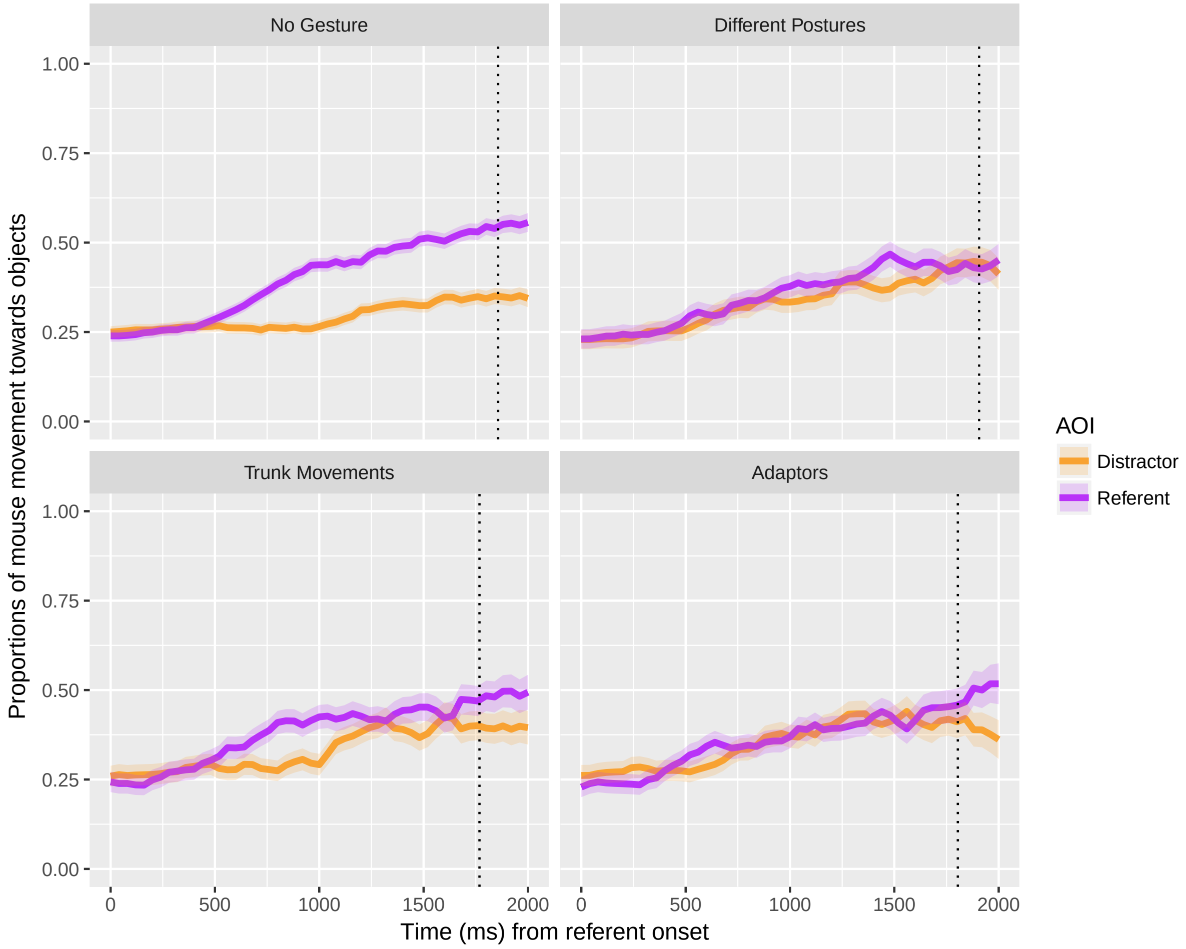

Figure 4. Mouse-tracking results for Experiment 2: Proportion of cumulative distance traveled toward each object from 0 to 2000 ms post-referent onset. Proportions were calculated from the total cumulative distance participants moved the mouse until that time bin (from video onset, when cursor was made visible). Shaded areas represent ± 1 standard error of the mean.

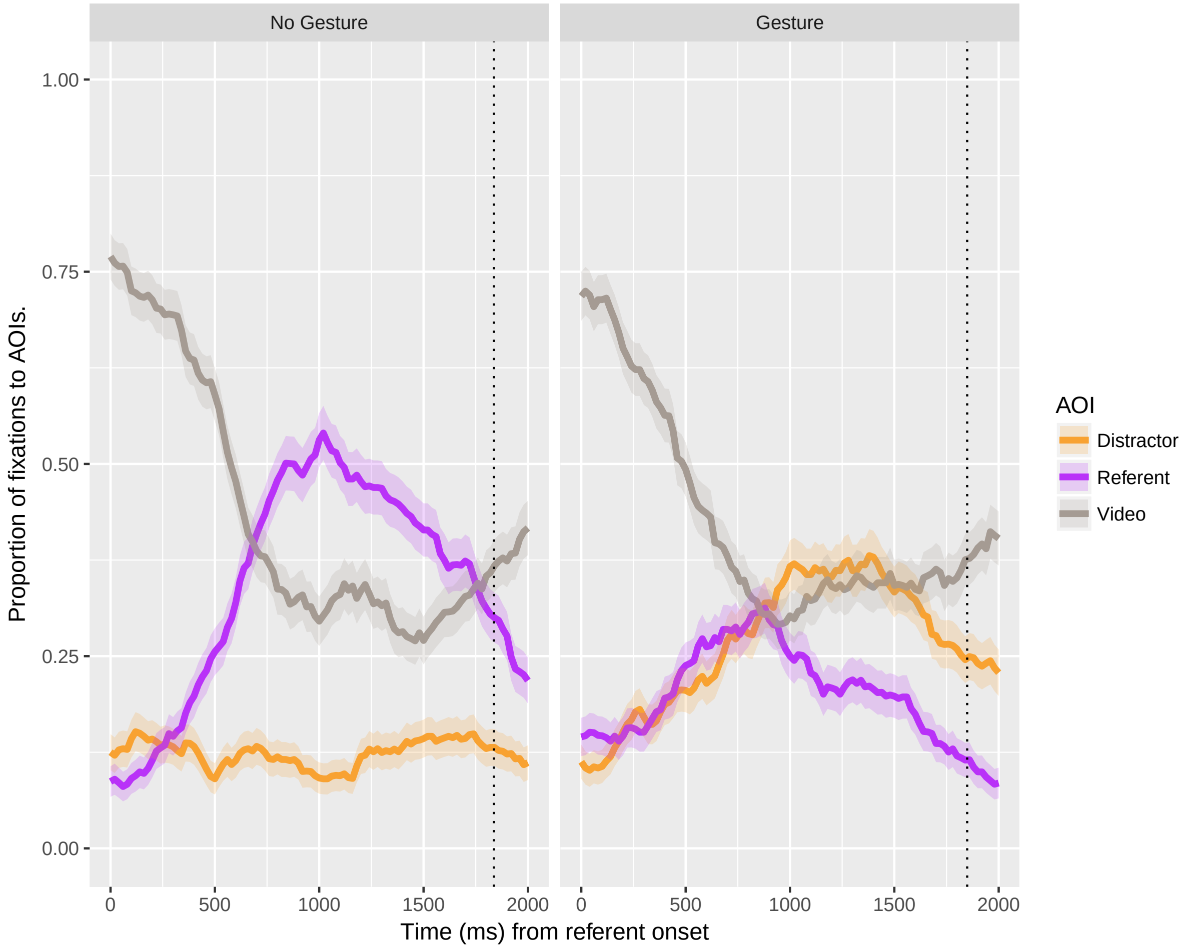

Figure 5. Eye-tracking results for Experiment 2: Proportion of fixations to each object (referent, distractor, video), from 0 to 2000 ms post-referent onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ± 1 standard error of the mean.

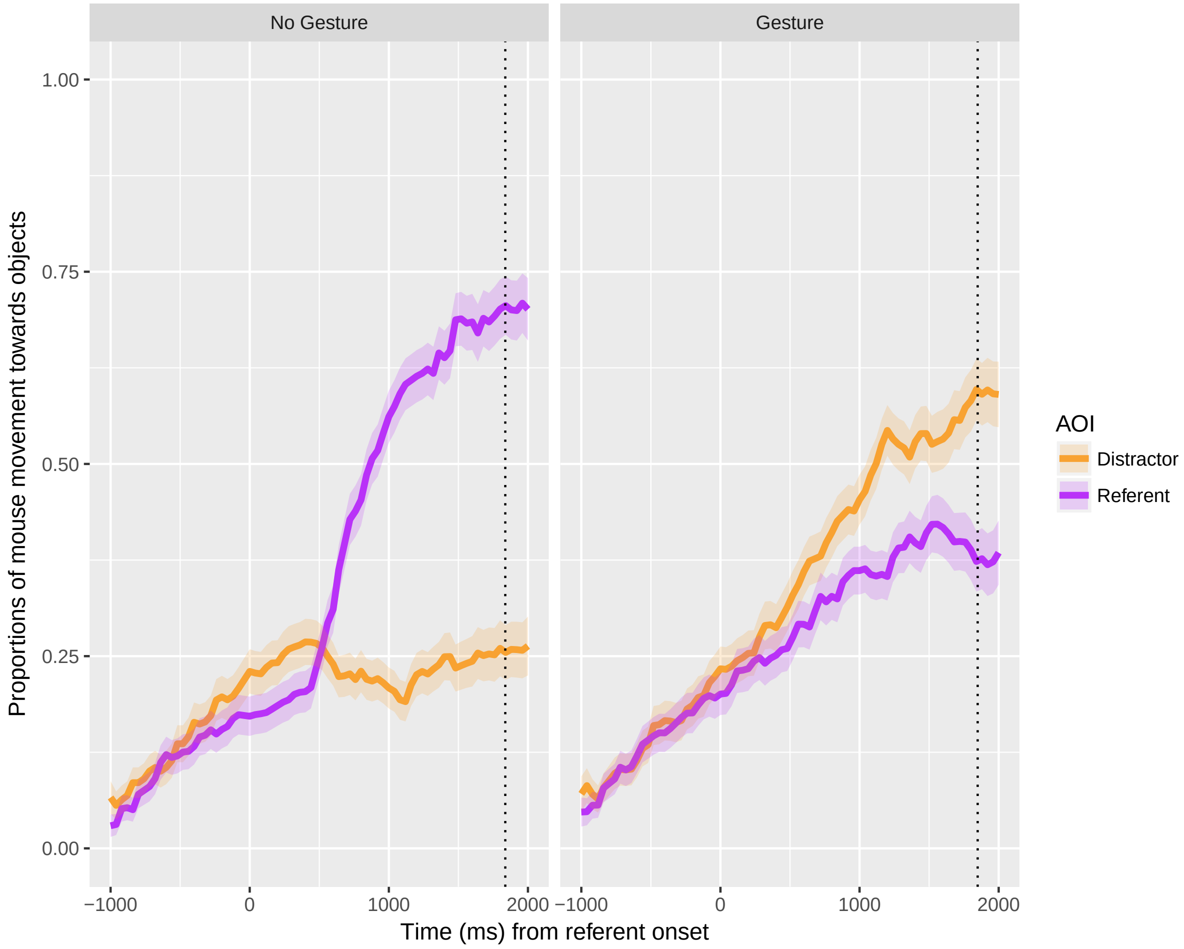

Figure 6. Mouse-tracking results for Experiment 2: Proportion of cumulative distance traveled toward each object from 0 to 2000 ms post-referent onset. Proportions were calculated from the total cumulative distance participants moved the mouse until that time bin (from speech-onset, when cursor was made visible). Shaded areas represent ± 1 standard error of the mean.